



Alma Mater Studiorum - Università di Bologna

Scuola di Dottorato in Scienze Economiche e Statistiche

Dottorato di Ricerca in

Metodologia Statistica per la Ricerca Scientifica

XXI ciclo

The Wrapping Approach for Circular Data Bayesian Modeling

Clarissa Ferrari

Dipartimento di Scienze Statistiche "P. Fortunati"

Marzo 2009



Alma Mater Studiorum - Università di Bologna

Scuola di Dottorato in Scienze Economiche e Statistiche

Dottorato di Ricerca in

Metodologia Statistica per la Ricerca Scientifica

XXI ciclo

The Wrapping Approach for Circular Data Bayesian Modeling

Clarissa Ferrari

Coordinatore

Prof.ssa Daniela Cocchi

Relatore

Prof.ssa Daniela Cocchi

Settore Disciplinare

SECS-S/01

Dipartimento di Scienze Statistiche "P. Fortunati"

Marzo 2009

As on the circle, the conclusion is not else than a new beginning.

A Marina (l'incipit iniziale),
me (la perseveranza in itinere),
Michela (il costante incoraggiamento e fondamentale supporto finale).

Abstract

Although circular data are special, they arise in many different contexts. Examples are found in earth sciences, meteorology, biology, physics, etc. Standard statistical techniques cannot be used to analyze circular data because of circular geometry of the sample space. There are different approaches to handle circular data. In the *embedding approach* the direction are treated as angles, while in the most popular *intrinsic approach* the direction are treated as unit complex number and modeled by von Mises distribution. An alternative, and more general class of distribution models can be obtained using the so called *wrapping approach*, in which the circular distributions are obtained wrapping the distributions on the real line onto the unit circle.

In this thesis, after giving a general overview about circular data, we deeply analyze the wrapping approach showing the main drawback and advantages of this method. Focusing on wrapped Normal distribution, we provide an approximation for this circular distribution that turns out to be very useful to improve the inferential results. This approximation, in fact, is directly used into the Bayesian inference procedure allowing to overcome the main disadvantage, the identifiability problem, and to show the flexibility and ease of applicability of this approach in model with complex structure as measurement error model and high dimensional spatial and spatiotemporal model. The main contribution of this work is substantially of overcoming the identifiability problem with the consequently possibility to apply the standard in line inferential procedures and methods to circular data as well.

In order to appreciate the flexibility and the ease of applicability and interpretability of the wrapping approach two original applications of measurement error model for circular data are presented: the first in a spatial context and the second in a dynamic spatiotemporal context. Some remarks and discussions about future developments conclude the thesis.

Acknowledgements

I would like to express my gratitude and appreciation to my (many) advisors: Prof.ssa Daniela Cocchi for her patience and encouragement, Prof. Alan Gelfand for his fundamental intuitions and suggestions and Prof.ssa Giovanna Jona Lasinio for her help and for having showed me “the way”.

Thanks to my PhD colleagues Davide, Massimo, Ida, Christian, Anna, Elena and Maria Serena for their cooperation and cheerful solidarity.

I would like to thank my overseas colleagues Gavino and Saki (for having taught me how to survive in a foreign country!) and Veronica. In the same way, I thank Marta for her invaluable bibliographic references.

A thanks and a big hug to Rosalba and Stefania (Suffy) for their “presence”.

Particular thanks to Sabrina, for having listened to me in difficult moments, and to Michela (from “ironhill”) for having shared with me the PhD adventure right from the beginning.

Finally, I express my gratitude to my parents for their support in all these years.

Thanks of all to everyone.

Contents

Abstract	iii
Acknowledgements	v
1 Introduction	1
1.1 Motivating Example	1
1.2 Examples of Circular Data and Graphical Representation	2
1.3 A Brief Tour	3
2 Circular Data	7
2.1 Introduction	7
2.2 Descriptive Statistics	8
2.2.1 Measures of location and concentration	8
2.2.2 Sample trigonometric moment	10
2.3 Circular Probability Distributions	11
2.3.1 Distribution function and characteristic function	11
2.3.2 Trigonometric moments and other population characteristics .	13
2.4 Statistical Approaches to Model Circular Data	14
2.4.1 The embedding approach	14
2.4.2 The intrinsic approach	15
2.4.3 Wrapping approach	19
2.4.4 Relationship among approaches and circular distributions . . .	23
2.5 Discussion	24
3 Wrapping Approach	27
3.1 Parameter Estimation	29
3.1.1 Simulated examples	31
3.2 Sample Size Effect on Parameter Estimation	35
3.3 Variability Effect on Parameter Estimation	40
3.4 Identifiability Problem	45
3.4.1 Wrapped density approximation	45
3.4.2 Tests of uniformity and threshold value for variance	49
3.5 Discussion	56

4	Measurement Error Model	59
4.1	ME Models Definition	59
4.2	Properties and Parameter Estimation Methods for ME Models . . .	61
4.3	ME models as a Missing Data Problem: the Bayesian Approach . . .	63
4.3.1	Bayesian methods when χ variable is unobservable	64
4.4	Wrapped ME model	65
4.5	Relationship between in line and Wrapped ME Model	67
4.6	Discussion	68
5	Spatial Modeling	75
5.1	Introduction to Spatial Models	75
5.2	Point-referenced Data Models	76
5.2.1	Stationarity	76
5.2.2	Variogram	77
5.2.3	Isotropy	78
5.2.4	Variogram model fitting	80
5.3	Hierarchical Modeling for Univariate Point-referenced Data	81
5.3.1	Spatial model definition	82
5.3.2	Hierarchical Bayesian methods	82
5.4	Wrapped Spatial Models	83
5.4.1	Examples	84
5.5	Discussion	90
6	Spatiotemporal Modeling	93
6.1	Spatiotemporal model formulation	93
6.2	Spatiotemporal Bayesian model specification	95
6.3	Dynamic Spatiotemporal modeling	96
6.3.1	Dynamic linear model	96
6.3.2	Updating equation	98
6.3.3	Dynamic formulation for spatiotemporal models	99
6.4	Wrapped Spatiotemporal Modeling	100
6.4.1	Simulated Example	102
6.5	Discussion	105
	Conclusions and Future Works	107
	Appendix	111
	A Simulation plan	111
	Bibliography	119

List of Figures

1.1	Fourteen buoys of the Italian wave network (RON).	2
1.2	Circular raw plots: (a) row plot of turtles data; (b) row plot and density estimate of British mallards data.	3
1.3	Rose diagrams: (a) rose diagram of mortality data; (b) wind rose diagram.	4
2.1	Representation of the direction x by angle θ and by complex number z	8
2.2	Density plot of circular Uniform distribution	16
2.3	von Mises distribution with concentration parameter equal to 0.5, 1, 2, 4.	18
2.4	In line Normal distribution (top panel) and corresponding wrapped Normal distribution (bottom panel).	23
3.1	Univariate example 1: histogram of the in line data Y , simulated from $N(0, \pi/4)$, and histogram and rose diagram of the corresponding wrapped data.	31
3.2	Univariate example 2: histogram of in line data Y , simulated from $N(1, \pi)$, and histogram and rose diagram of the corresponding wrapped data.	32
3.3	Traces of the estimated parameter chains of univariate examples 1 and 2	32
3.4	Bivariate example: scatter plot of the in line data Y , simulated from $N_2(\mu = (0, 1)', \Sigma)$, and histogram and rose diagram of the corresponding wrapped components.	33
3.5	Autocorrelation functions of the estimated parameters of the univariate example 1 and 2.	35
3.6	Histograms of the in line data simulated from $N(0, \pi/4)$ and corresponding rose diagrams, density estimates (red lines) and means (blue dots) of the wrapped data. The number of replications n is respectively equal to 10 (a), 40 (b), 100 (c) and 300 (d).	36
3.7	MCMC chain traces for μ and σ^2 parameters of the four simulated data sets. Outputs obtained with flat priors $\mu \sim N(0, 25)$ and $\sigma^2 \sim \text{InvGamma}(4, 8)$	38

3.8	MCMC chain traces for μ and σ^2 parameters of the four simulated data sets. Outputs obtained with informative priors $\mu \sim N(0, 2)$ and $\sigma^2 \sim \text{InvGamma}(4, 2)$	39
3.9	Histograms of the in line data simulated from $N(0, \sigma^2)$ and corresponding rose diagrams, density estimates (red lines) and means (blue dots) of wrapped Normal data. The variance σ^2 is respectively equal to $\pi/10$ (a), $\pi/8$ (b), $\pi/6$ (c), $\pi/4$ (d), $\pi/2$ (e), π (f) and $3\pi/2$ (g). . .	41
3.10	MCMC chain traces for μ and σ^2 parameters of the seven simulated data sets. Outputs obtained with flat priors $\mu \sim N(0, 30)$ and $\sigma^2 \sim \text{InvGamma}(4, 8)$	43
3.11	MCMC chain traces for μ and σ^2 parameters of the seven simulated data sets. Outputs obtained with informative priors: $\mu \sim N(0, 1)$ and $\sigma^2 \sim \text{InvGamma}(4, 2)$	44
3.12	Normal density and its corresponding Wrapped Normal density with $\mu = \frac{\pi}{2}$ (left) and $\mu = \frac{\pi}{2} + 2\pi$ (right)	45
3.13	Probability density plots of Normal distributions conditioned on wrapped values $X = x$, related to Tables 3.10-3.13	48
3.14	Plots of wrapped Normal distributions (n=100) with different variances.	53
4.1	Plot of simulated in line distributions (left plots) and corresponding wrapped distributions (right rose diagram plots) inherent to distributions reported in Table 4.1	71
4.2	Plot of simulated in line distributions (left plots) and corresponding wrapped distributions (right rose diagram plots) inherent to distributions reported in Table 4.2	72
5.1	Variogram and its parameters.	78
5.2	Theoretical semivariogram: (a) linear, (b) spherical, (c) exponential.	79
5.3	Traces of estimated parameters of Spatial WN Process with dimensions d=3	86
5.4	Traces of estimated parameters of Spatial WN Process with dimensions d=6	86
5.5	Map of the 27 spatial location in the Adriatic sea	88
5.6	Observed values, estimated values and 95% posterior interval for the parameter vector μ	89
6.1	Traces of the estimate parameters of wrapped spatiotemporal dynamic model.	105

List of Tables

3.1	Estimates of the two univariate wrapped Normal distributions plotted in Figure 3.1	31
3.2	Parameter estimates of the bivariate wrapped Normal distribution plotted in Figure 3.4	33
3.3	Geweke Z-score diagnostic for the parameters of both univariate examples.	34
3.4	Geweke Z-score diagnostic (computed with burn-in of 1000 iteration) for the chains depicted in Figures 3.7 and 3.8	37
3.5	Posterior results for the four data sets plotted in Figure 3.6, obtained using the flat priors: $\mu \sim N(0, 25)$ and $\sigma^2 \sim InvGamma(4, 8)$	37
3.6	Posterior results for the four data sets plotted in Figure 3.6, obtained using the priors: $\mu \sim N(0, 2)$ and $\sigma^2 \sim InvGamma(4, 2)$	38
3.7	Posterior results for the seven data sets plotted in Figure 3.7, obtained using the following flat priors: $\mu \sim N(0, 30)$ and $\sigma^2 \sim InvGamma(4, 8)$	40
3.8	Posterior results for the seven data sets plotted in Figure 3.7, obtained using the following informative priors: $\mu \sim N(0, 1)$ and $\sigma^2 \sim InvGamma(4, 2)$	42
3.9	Geweke Z-score diagnostic (computed with burn-in of 5000 iterations) for the chains depicted in Figures 3.10 and 3.11	42
3.10	Conditional probabilities using $Y \sim N(0, \pi/2)$	46
3.11	Conditional probabilities using $Y \sim N(0, \pi)$	47
3.12	Conditional probabilities using $Y \sim N(0, 4\pi/3)$	47
3.13	Conditional probabilities using $Y \sim N(0, 2\pi)$	47
3.14	Probability that the wrapping has no effect ($Pr(Y = x)$) related to four Normal distributions with zero means and different variances listed in Tables 3.10-3.13	48
3.15	Tests of uniformity for wrapped Normal density with sample size $n = 30$. The bold line indicates the first variance value beyond which the test accepts the hypothesis of uniformity.	52
3.16	Tests of Uniformity for wrapped Normal density with sample size $n = 100$. The bold line indicates the first variance value beyond which the test accepts the hypothesis of uniformity.	52
3.17	Tests of Uniformity for wrapped Normal density with sample size $n = 1000$. The bold line indicates the first variance value beyond which the test accepts the hypothesis of uniformity.	53

3.18	Numerically approximations of $A(h)$ function and corresponding variances.	55
4.1	Table of simulated in line distributions and corresponding wrapped distributions. In the A, B and C cases the unobserved variable variance is fixed while in the D, E and F cases the fixed parameter is the error variance.	69
4.2	Table of simulated in line distributions and corresponding wrapped distributions. The variance ratio $\sigma_{\xi}^2/\sigma_{Y_{true}}^2$ is constant.	70
4.3	ME model variance estimates	73
5.1	Semivariograms for common parametric isotropic models.	80
5.2	Covariance functions for common parametric isotropic models.	81
5.3	Parameter estimates of the wrapped spatial Normal processes with dimension $d = 3$	85
5.4	Parameter estimates of the wrapped spatial Normal processes with dimension $d = 6$	85
5.5	Simulation plan scheme for wrapped spatial process with dimension $d = 3$	91
5.6	Observed values, posterior estimates and posterior interval of the wave direction (expressed in radiant) on 25km x 25km lattice (UTM coordinates) recorded in a portion of Adriatic sea between Ancona and Ortona. The points identified by "o" are related to Ortona coast, while the points with "a" are related to Ancona coast.	92
6.1	Parameter estimates of the wrapped spatiotemporal dynamic model.	104
6.2	Geweke Z-score diagnostic for the parameters of wrapped spatiotemporal dynamic model.	104
A.1	Simulation plan scheme for wrapped spatial process with dimension $d = 6$	112
A.2	Simulation plan scheme for wrapped spatial process with dimension $d = 15$	113
A.3	Simulation plan results for wrapped spatial process with dimension $d = 3$ (first part)	114
A.4	Simulation plan results for wrapped spatial process with dimension $d = 3$ (second part)	115
A.5	Simulation plan results for wrapped spatial process with dimension $d = 6$ (first part)	116
A.6	Simulation plan results for wrapped spatial process with dimension $d = 6$ (second part)	117

Chapter 1

Introduction

Circular data arise in many different contexts. Many examples of this kind of data are found in earth science, meteorology (as wind direction analysis), biology (e.g. study of animal movement direction), physics and, more in general, in any context where the study of data recorded in degrees or radius in a circle is needed. Although it is quite common to find and deal with circular data, their handling arises several difficulties. Their particularity is due to the support, finite and in the unit circle $[0, 2\pi)$, and to the dependence of descriptive and inferential results on choosing the starting point in the unit circle. These features make this kind of data special and different from *in line* data, so that ad hoc statistical methods and modeling are needed for their management. Henceforth, we refer to data with support in \mathbb{R} or in a subset of \mathbb{R} as ‘*in line*’ data whereas for data with support in $[0, 2\pi)$ we specify the adjective ‘*circular*’.

In the following section we introduce the motivating example, regarding a meteorological-marine issue, that leads us to deal with circular data; whereas in Section 1.2 we give some examples and graphical representations of these data. Finally, we conclude this chapter providing an essential review of circular data literature and describing the structure of the thesis.

1.1 Motivating Example

Most of the studies carried out on marine data are based on outputs from deterministic models, usually climatic forecasts computed at several spatial and temporal resolution. Wave heights and wave directions, these last being circular data measured in degrees, are the main results of marine forecasts. The principal provider for global numerical wave forecasts in Europe is the European Center for Medium-Range Weather Forecasts (ECMWF), which runs as global medium range (3-5 up to 10 day forecasts, 55 km spatial resolution) and as high resolution short term (3 days, 25 km resolution Wave Amplitude Model, *WAM*) models in the Mediterranean Area. Despite all efforts and improvements in model implementation and the growing amount of assimilated data, wave heights and wave directions are often biased and not efficiently estimated by numerical models. It arises, thus, the need to improve the quality of the forecasting system by a statistical post-processing that aims

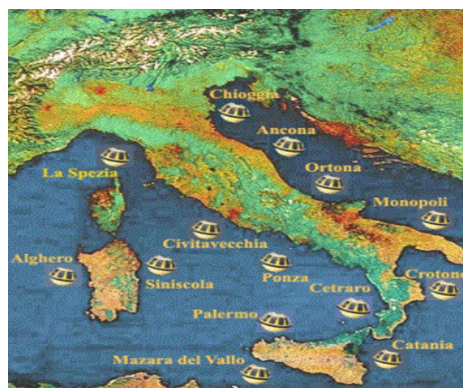


Figure 1.1: Fourteen buoys of the Italian wave network (RON).

to correct the estimates through observed data given by the Italian wave network (RON) constituted of 14 buoys (see Figure 1.1).

A first tentative to calibrate the computed data is founded in Bruschi *et al.* (2005), but in that work the difficulty to deal with circular data leads directly the authors to choose for a simplified semiparametric model where the main directions of the wave (i.e. the circular data) occur only as a dummy variables. The starting motive of this thesis is, then, to find a flexible and easily applicable method to manage and model directly circular data. Of course, it exists a quite wide literature about circular data (see e.g. Mardia (1972), Mardia and Jupp (1999), Jammalamadaka and SenGupta (2001) or Fisher (1993)) but it is limited to descriptive statistics and simple and standard models as the regression model or simple temporal series models.

Our goal, here, is to find out a flexible and easily interpretable procedure that allows to extend to circular data all the models and inferential procedures applied to in line data, including the fairly complicated model such as the spatial and spatiotemporal ones.

1.2 Examples of Circular Data and Graphical Representation

The oldest observed circular data sets regard the studies of animal navigation. Famous are the turtle data about animal orientation after laying eggs cited in Stephens (1963). Other examples of circular data in biology are the mallard data of the British Ornithologists' Union and the swimming directions of the *Daphnia* (Waterman and Jander's data cited in Waterman (1963)).

In the context of image analysis, circular data occur, for example, in machine vision (Mardia *et al.*, 1996) or in orientation of textures (Blake and Marinos, 1990).

Also in the medicine context it is possible to find circular data as, for instance, the incidence of onsets of a particular disease (or of deaths) at various times of the year. Perhaps, the more profitable context where circular data have been studied and used is surely in physics. Von Mises (1918) introduced his famous probability distribution in order to study the deviation of measured atomic weights from integral values.

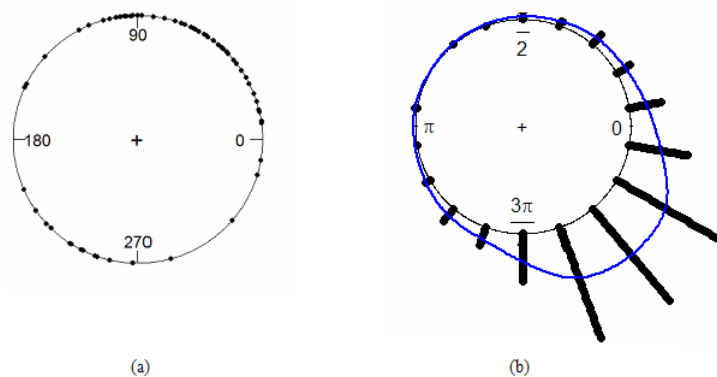


Figure 1.2: Circular raw plots: (a) row plot of turtles data; (b) row plot and density estimate of British mallards data.

Generally speaking, we can assert that we are able to find circular data in many other contexts and environments and as many graphical representation are available to depict circular data, as well. Some of them are here presented.

The easiest plot is the circular raw plot, where the data are drawn directly on the circle. For instance, the raw circular plots for turtle data and mallard data are depicted in Figure 1.2.

Sometimes it can be useful to borrow circular plots from the in line graphical representation in order to understand the difference or similarity between in line and circular data. The corresponding plot of the in line histogram is given by the rose diagram in which the bars of the histogram are replaced by sectors. Two examples are given in Figure 1.3.

1.3 A Brief Tour

Perhaps the oldest reference about circular data goes back to 1918 when von Mises introduced his probability density. Besides von Mises, between the authors that have given the major contribute to circular statistics we can include Watson and Stephens with their results about goodness of fit and statistic test theory (Watson, 1961; Stephens, 1963, 1970). Between the 60's and the 80's, Kent contributed to the study of complex circular distributions and one of his more important work is reported in Kent (1978). More recently, in Fisher (1993) we find a comprehensive treatise on circular data distributions with particular attention to non-parametric methods in spatial structure analysis. Finally, in Mardia (1972) and in Mardia and Jupp (1999) there are the more important results and knowledge about approaches, probability distribution theory and inference for circular data. But only in the last years, by the developing of more computationally efficient estimation procedure such as MCMC methods and the EM algorithm, statistics for circular data has regarded, other than the descriptive and the probability distribution analysis, also the computing of simple linear models (see Harrison and Kanji, 1988; Fisher, 1993; Fisher and Lee, 1992), linear models in a Bayesian context (Guttorp and Lockhart,

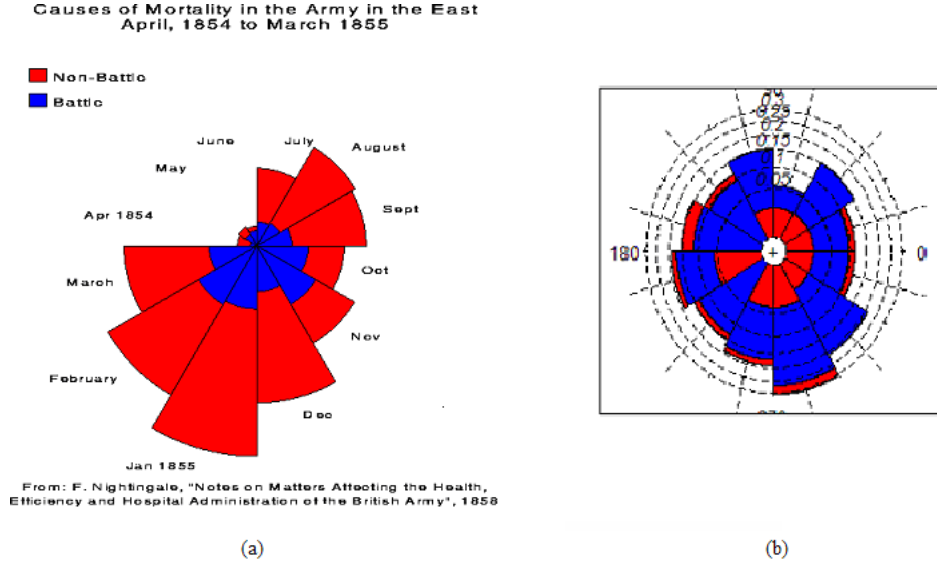


Figure 1.3: Rose diagrams: (a) rose diagram of mortality data; (b) wind rose diagram.

1988; Damien and Walker, 1999) and models for circular time series (see Breckling, 1989; Coles, 1998; Mardia and Jupp, 1999; Ravindran, 2002; Hughes, 2007).

This is only a short review of the major authors and works about circular data analysis but it is sufficient to realize that, probably due to the difficulties peculiar to this kind of data, their modeling is limited to quite simple and low dimensional models. As a matter of fact, there are not in literature examples of high dimensional circular data models or circular data analysis in spatial or spatiotemporal structure or, more in general, circular data models having complex structure. The main goal of this thesis is to provide a flexible and of ease applicability and interpretability modeling tool for circular data. In particular, our aim is to find out a procedure that allows to overcome the problems and difficulties due to the peculiar features to circular data in order to extend to this kind of data all the models and methods already employed for standard *in line* statistical analysis.

The main contribution of this work is substantially methodological and the presented applications regard basically simulated data, even if a real data application is provided in Chapter 5 as an example to appreciate the goodness of inference results. In order to have a general overview about circular data and their main features and properties, in Chapter 2 we provide some descriptive statistics and describe the three approaches (embedding, intrinsic and wrapping) for handling circular data. For each of them, we present the main circular distributions describing the probability function and properties. Moreover, the relationship between circular distributions and a comparison in terms of advantages and drawbacks of the three approaches is discussed.

In Chapter 3 we deeply analyze the wrapping approach and demonstrate the reason why we consider this approach the most suitable in the pursuit of our goal. Following the work of Coles (1998), we describe the parameter estimation procedure for

the wrapped Normal distribution, giving also some simulated examples. Then, we analyze the effect of sample size and data variability on inference results, revealing the main drawback of the wrapping approach: the identifiability problem concerning the wrapping coefficients. The main contribution of this thesis consists in solving the wrapping coefficients identifiability problem demonstrating that all the wrapped Normal distribution can be approximated adequately by only the three wrapping coefficient values $\{-1, 0, 1\}$. This evidence, which is originally showed here, is directly used into the Bayesian inference procedure through an opportune prior probability on the k coefficients, obtaining a considerable improvement of the inference results. This new setting of the inference procedure allows the computing of models as the high dimensional multivariate ones and models with complex structure as the measurement error, spatial and spatiotemporal models.

In Chapter 4, we present a definition and a general exposition of measurement error (ME) models and successively we show the ease of derivation of the ME wrapped model. Then, in order to verify the relationship between in line and wrapped ME model, a simulation-based study is carried out in which particular attention is dedicated to sensitivity analysis to the error variance on parameter estimation.

The first interesting application of the Bayesian inference procedure derived in Chapter 3 regards a circular spatial model implemented in Chapter 5. For this purpose we first give a general overview of the in line spatial modeling and then we focus on point-referenced data models. For this kind of spatial models, we illustrate a hierarchical Bayesian modeling that constitutes the basic structure even in case of wrapped spatial model, which is successively derived. It is worth to note that the derivation of the circular spatial model is particularly easy and intuitive using the wrapping approach. Considering the spatial process as a characterization of a multivariate circular distribution on d spatial locations, the wrapped spatial process is directly obtained applying the wrapping procedure described in Chapter 3 to each component of the process. This peculiar feature to wrapping approach, together with the easy interpretability of the wrapping distribution parameters, are the strength of this method. In fact, the easy extension of the wrapped model to the spatial context is possible thanks to the straightforward modeling of the multivariate circular distributions. In order to show the flexibility of this approach, we show some simulated examples inherent two spatial processes with different dimensions. Moreover, an extensive simulation plan is projected to find out the better parameter set up in case of high dimensional multivariate model. The case of a spatial process for real wave direction data recorded in 27 spatial locations in Adriatic sea concludes the chapter.

The application where the flexibility of the wrapping approach appears more evidently and, at the same time, essential for the implementation of complex models, is reported in Chapter 6. In this chapter, in fact, we introduce an in line spatiotemporal model and its related Bayesian model specification. Moreover, a dynamic spatiotemporal modeling is illustrated. Even in this case of a very complex model structure, we can derive the corresponding wrapped circular model straightforwardly. A simulated example of a dynamic spatiotemporal measurement error model concludes the chapter. Even and especially in this case, where the complexity of the implemented

model induces an inference procedure computationally intensive, the implementation is made possible thanks to the ease ‘transportability’ of the in line inference procedure to the circular wrapped model.

It is worth to note that the models presented in Chapters 5 and 6 represent the first examples in literature of spatial (point-referenced) and dynamic spatiotemporal models, respectively, for circular data.

Chapter 2

Circular Data

2.1 Introduction

Circular data are rather special, but they arise in many different contexts. Many examples of circular data are found in various scientific fields such as earth sciences, meteorology (as for example the wind directions), biology (e.g. the orientations of turtles after laying eggs), physics, etc. Standard statistical techniques can not be used to analyze circular data. This is due to the circular geometry of the sample space. For example, let x_1, x_2, \dots, x_n be independent observations on the unit circle, such that $0 < x_i < 2\pi, i = 1, 2, \dots, n$; the mean direction \bar{x} , that is the mean of the x_1, \dots, x_n circular observations, is not equal to $\frac{1}{n} \sum_{i=1}^n x_i$. To see this, consider a sample of size 2 on the circle consisting of the angles 1° and 359° . Cutting the circle at 0° would give the sample mean direction as 180° , whereas cutting the circle at $180^\circ = (-180^\circ)$ would give the sample mean as 0° . So, the resulting summary statistics depend strongly on the point where the circle is cut.

It turns out that the appropriate way of constructing summary statistics for circular data is to regard points on the circle as unit vectors in the plane and then take polar coordinates of the sample mean of these vectors.

There are two useful ways of regarding directions in the plane: as angles and as unit complex numbers. So, chosen an initial direction and orientation for the unit circle, each point x on the circle can be represented by an angle θ or equivalently by a unit complex number z . These quantities are related to x by

$$x = (\cos \theta, \sin \theta)^T \text{ and } z = e^{i\theta} = \cos \theta + i \sin \theta,$$

as shown in Figure 2.1

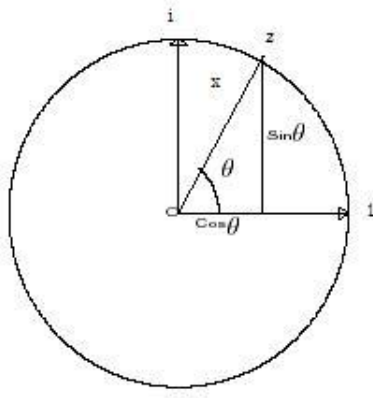


Figure 2.1: Representation of the direction x by angle θ and by complex number z .

These two ways to represent directions can be regarded respectively as an *intrinsic approach* (directions considered as points on the circle itself) and *embedding approach* (directions are special points in the plane).

More details about these approaches will be argued in Section 2.4.

A last notation is useful especially in the next chapters in order to avoid misunderstanding.

2.2 Descriptive Statistics

In this section we summarize some circular descriptive statistics useful to introduce the argument and to understand the problems related to using this kind of data. In particular, we describe the main location and dispersion quantities.

2.2.1 Measures of location and concentration

Let x_1, \dots, x_n be unit vectors with corresponding angles θ_i , $i = 1, \dots, n$. The *mean direction* $\bar{\theta}$ of $\theta_1, \dots, \theta_n$ is the direction of the resultant $x_1 + \dots + x_n$ of x_1, \dots, x_n . Since the Cartesian coordinates of x_i are $(\cos \theta_i, \sin \theta_i)$, for $i = 1, \dots, n$, we can define

$$\bar{C} = \frac{1}{n} \sum_{i=1}^n \cos \theta_i, \quad \bar{S} = \frac{1}{n} \sum_{i=1}^n \sin \theta_i \quad (2.1)$$

Therefore, the mean direction, $\bar{\theta}$, is the solution of the equations:

$$\cos \bar{\theta} = \bar{C} / \bar{R}, \quad \sin \bar{\theta} = \bar{S} / \bar{R} \quad (2.2)$$

where the *mean resultant length* \bar{R} , associated with the *mean direction* $\bar{\theta}$, is given by

$$\bar{R} = (\bar{C}^2 + \bar{S}^2)^{1/2}. \quad (2.3)$$

When $\bar{R} > 0$, $\bar{\theta}$ can be written explicitly as follows:

$$\bar{\theta} = \begin{cases} \tan^{-1}(\bar{S}/\bar{C}) & \text{if } \bar{C} \geq 0 \\ \tan^{-1}(\bar{S}/\bar{C}) + \pi & \text{if } \bar{C} < 0 \end{cases} \quad (2.4)$$

From (2.1) and (2.2) it follows that

$$\bar{R} = \frac{1}{n} \sum_{i=1}^n \cos(\theta_i - \bar{\theta}) \quad (2.5)$$

and for $\bar{R} > 0$

$$\sum_{i=1}^n \sin(\theta_i - \bar{\theta}) = 0 \quad (2.6)$$

Equation (2.6) is analogous to

$$\sum_{i=1}^n (y_i - \bar{y}) = 0$$

where y_1, \dots, y_n are the in line observations with sample mean \bar{y} . As a matter of fact, equations (2.5) and (2.6) state that the sums of deviations about the mean are zero.

It is not difficult to show that the sample mean direction is *invariant* under rotation. This can be shown supposing a new choice of initial direction, making angle α with the original initial direction and rewrite the data points as $\theta'_i = \theta_i - \alpha$ and $\bar{C}' = \frac{1}{n} \sum_{i=1}^n \cos \theta'_i$, $\bar{S}' = \frac{1}{n} \sum_{i=1}^n \sin \theta'_i$. This equivariance is analogous to the equivariance under translation of the sample mean of the in line observations.

Another important measure of location is the median direction. In an analogous way for data on the line, it is possible to define the *sample median direction* $\tilde{\theta}$ of the angles $\theta_1, \dots, \theta_n$ as any angle ϕ such that: (i) half of the data points lie in the arc $[\phi, \phi + \pi)$, and (ii) the majority of the data points are nearer to ϕ than to $\phi + \pi$. In particular, when the sample size n is odd, the median is one of the data points, otherwise the sample median is taken as the midpoint of two appropriate adjacent data points.

Measures of Concentration and Dispersion

In order to analyze concentration and dispersion of circular data, mean resultant length \bar{R} given by (2.3) becomes very important. It lies in the range $[0, 1]$ and, in particular, $\bar{R} = 1$ implies that all the data points coincide. Thus \bar{R} is a measure of *concentration* of the data. However, $\bar{R} = 0$ does *not* imply uniform dispersion around the circle. To see this, we note, for example, that any data set of the form $\theta_1, \dots, \theta_n, \theta_1 + \pi, \dots, \theta_n + \pi$ has $\bar{R} = 0$ without being equally distributed around the circle.

The *resultant length* R is the length of the vector resultant $x_1 + \dots + x_n$, thus,

$$R = n\bar{R}$$

. For most descriptive and inferential purposes, the mean resultant length \bar{R} is more important than any measure of dispersion. Using of \bar{R} in an inferential purposes is argued in Section 3.4.

Moreover, using \bar{R} it is possible to define the *circular variance*:

$$V = 1 - \bar{R} \quad (2.7)$$

Similarly to in line variance, the smaller the value of the circular variance is, the more concentrated the distribution will be. However, note that $0 \leq V \leq 1$, unlike an ordinary in line variance. Also, note that in the light of the above remark concerning the interpretation of $\bar{R} = 0$, $V = 1$ does not necessarily imply a maximally dispersed distribution.

Other measures of dispersion are the *sample circular standard deviation*, given by

$$v = [-2 \log(1 - V)]^{1/2} = [-2 \log \bar{R}]^{1/2} \quad (2.8)$$

and the *sample circular dispersion*:

$$\hat{\delta} = \frac{1 - \bar{R}_2}{2\bar{R}^2} \quad (2.9)$$

where \bar{R}_2 denotes the mean resultant length of the doubled angles $2\theta_1, \dots, 2\theta_n$. The circular dispersion $\hat{\delta}$ plays an important role in calculating a confidence interval for a mean direction and in comparing and combining several sample mean directions (see Fisher (1993, §4.4.4, §5.4.2)).

2.2.2 Sample trigonometric moment

In Section 2.2.1 the moments \bar{C} and \bar{S} were introduced for defining the sample mean direction and the sample circular variance. It is useful to combine them into the first trigonometric moment. In particular, $\bar{\theta}$ and \bar{R} are the angular and amplitude components of the *first trigonometric moment about the zero direction*

$$m_1' = \bar{C} + i\bar{S} = \bar{R}e^{i\bar{\theta}} \quad (2.10)$$

Extending this notation, we can define the p^{th} *trigonometric moment about the zero direction* for $p = 1, 2, \dots$ as

$$m_p' = \bar{C}_p + i\bar{S}_p = \bar{R}_p e^{i\bar{\theta}_p} \quad (2.11)$$

where

$$\bar{C}_p = \frac{1}{n} \sum_{i=1}^n \cos p\theta_i \quad \bar{S}_p = \frac{1}{n} \sum_{i=1}^n \sin p\theta_i \quad (2.12)$$

and $\bar{\theta}_p$ and \bar{R}_p denote the sample mean direction and sample mean resultant length of $p\theta_1, \dots, p\theta_n$. The p^{th} *sample trigonometric moments about the mean direction* are obtained in a similar way:

$$m_p = \frac{1}{n} \sum_{i=1}^n \cos p(\theta_i - \bar{\theta}) + i \frac{1}{n} \sum_{i=1}^n \sin p(\theta_i - \bar{\theta}) \quad (2.13)$$

In particular, using equation (2.5) and (2.6) it holds that

$$m_1 = \frac{1}{n} \sum_{i=1}^n \cos(\theta_i - \bar{\theta}) = \bar{R} \quad (2.14)$$

(because $\sum_{i=1}^n \sin(\theta_i - \bar{\theta}) = 0$), and

$$m_2 = \frac{1}{n} \sum_{i=1}^n \cos 2(\theta_i - \bar{\theta}) \quad (2.15)$$

From the first and second trigonometric moments about zero direction, m_1' and m_2' , we obtain measures of *skewness* and *kurtosis* for circular data defined, respectively, as

$$\hat{s} = \frac{\bar{R}_2 \sin(\bar{\theta}_2 - 2\bar{\theta})}{(1 - \bar{R})^{3/2}} \quad (2.16)$$

and

$$\hat{k} = \frac{\bar{R}_2 \cos(\bar{\theta}_2 - 2\bar{\theta}) - \bar{R}^4}{(1 - \bar{R})^2} \quad (2.17)$$

For symmetric unimodal data sets, \hat{s} is nearly zero. Data from an unimodal distribution, tend to have a sample kurtosis value around zero, while more peaked distributions have positive sample kurtosis.

Other measures of location and spread

Can be useful to define measures as the mode and the range for circular data. The *sample modal direction* $\check{\theta}$ is the direction corresponding to the maximum concentration of the data. One way of determining $\check{\theta}$ is to find the value of θ maximizing a density estimate $f(\theta)$, although this method depends on the amount of smoothing involved in the density estimation procedure (see Fisher (1993, §2.2))

Finally, turning to measures of spread, we define the *sample range* as the length of the smallest arc which contains all the observations.

2.3 Circular Probability Distributions

In this section basic concepts of distributional theory for circular data are introduced. For this purpose, the distribution function and characteristic function of circular data are defined. Moreover, population versions of trigonometric moments and measures of locations and dispersion are introduced in order to define the probability distribution and other characteristics of circular data.

2.3.1 Distribution function and characteristic function

Let $f(\theta)$ be the *probability density function* of a continuous random variable Θ , i.e. $f(\theta)$ is a non-negative 2π periodic function such that $f(\theta + 2\pi) = f(\theta)$ and

$\int_0^{2\pi} f(\theta)d\theta = 1$. The *distribution function* $F(\theta)$ can be defined over any interval (θ_1, θ_2) by $F(\theta_2) - F(\theta_1) = \int_{\theta_1}^{\theta_2} f(\theta)d\theta$. In particular, after choosing an initial direction and an orientation of the unit circle (generally 0° direction and anticlockwise orientation), $F(\theta)$ is defined as

$$F(\theta) = \int_0^\theta f(\phi)d\phi \quad (2.18)$$

obviously it holds that, $F(2\pi) = 1$.

Another useful tool for handling the distribution of a random angle θ is the function $p \mapsto E[e^{ip\theta}]$. Since θ and $\theta + 2\pi$ represent the same direction, it is necessary to restrict t to integer values only. The *characteristic function* of a random angle θ is the doubly-infinite sequence of complex numbers $\{\phi_p : p = 0, \pm 1, \dots\}$ given by

$$\phi_p = E[e^{ip\theta}] = \int_0^{2\pi} e^{ip\theta} f(\theta)d\theta, \quad p = 0, \pm 1, \pm 2, \dots \quad (2.19)$$

In many cases, it is useful to remember the following equivalence $e^{ip\theta} = \cos p\theta + i \sin p\theta$, for instance, in order to write the characteristic function, (2.19), as

$$\phi_p = \int_0^{2\pi} \cos p\theta f(\theta)d\theta + i \int_0^{2\pi} \sin p\theta f(\theta)d\theta \quad (2.20)$$

The complex numbers $\{\phi_p : p = 0, \pm 1, \dots\}$ are the Fourier coefficients of the distribution function $F(\theta)$. When ϕ_p 's are related to $F(\theta)$ by the formula (2.19), it is usual to approximate

$$f(\theta)d\theta \cong \frac{1}{2\pi} \sum_{p=-\infty}^{\infty} \phi_p e^{-ip\theta} \quad (2.21)$$

The relationship (2.21) does not carry any implication that the series is convergent, still less that it converges to $F(\theta)$. However, from the last equation and under opportune moment convergence constraint (i.e. if $\sum_{p=1}^{\infty} ((E[\cos p\theta])^2 + (E[\sin p\theta])^2)$ is convergent), the random variable Θ has a density $f(\theta)$ defined almost everywhere as

$$f(\theta) = \frac{1}{2\pi} \sum_{p=-\infty}^{\infty} \phi_p e^{-ip\theta} = \frac{1}{2\pi} \left\{ 1 + 2 \sum_{p=1}^{\infty} (\alpha_p \cos p\theta + \beta_p \sin p\theta) \right\} \quad (2.22)$$

The characteristic function has the following properties:

- i) a probability distribution on the circle is determined by its characteristic function (uniqueness property);
- ii) weak convergence of distributions is equivalent to point-wise convergence of characteristic functions, i.e. a sequence F_1, F_2, \dots of distribution functions converges weakly to F if and only if $\phi_p^{(n)} \rightarrow \phi_p$ for $p = 0, \pm 1, \dots$, where $\phi_p^{(n)}$ and ϕ_p denote the characteristic function of F_n and F (see Mardia (1972, §3.3, §4.2));

- iii) let Θ_1 and Θ_2 be two angular random variables; the characteristic function of (Θ_1, Θ_2) is defined as $\phi_{p,q} = E[e^{ip\theta_1+iq\theta_2}]$;
- iv) the variables Θ_1 and Θ_2 are independent if and only if $\phi_{p,q} = \phi_p \phi'_q$ (where ϕ and ϕ' are the marginal characteristic functions of Θ_1 and Θ_2);
- v) let $S_n = \Theta_1 + \dots + \Theta_n$ be the sum of n identically distributed random variables with common characteristic function ϕ_p , then the characteristic function of S_n is ϕ_p^n .

2.3.2 Trigonometric moments and other population characteristics

In analogy with the sample trigonometric moments, (2.10)-(2.13), we can easily define the *population* trigonometric moments.

The p^{th} trigonometric moment of $f(\theta)$, $p = 1, 2, \dots$, is given by

$$\begin{aligned} \mu'_p &\equiv \rho_p e^{i\mu'_p} = \rho_p \cos \mu'_p + i \sin \mu'_p \\ &= \int_0^{2\pi} \cos p\theta f(\theta) d\theta + i \int_0^{2\pi} \sin p\theta f(\theta) d\theta \end{aligned} \quad (2.23)$$

$$\mu'_p = \alpha'_p + i\beta'_p \quad (2.24)$$

where α'_p and β'_p are the p^{th} cosine and sine moments respectively. When $p = 1$, we write ρ for ρ_1 and μ for μ_1 , i.e.

$$\mu'_1 = \rho e^{i\mu} \quad (2.25)$$

where μ is the *mean direction* and ρ is the (*population version of*) *mean resultant length*.

The p^{th} central (centred about its mean direction μ) trigonometric moment of Θ is

$$\begin{aligned} \mu_p &\equiv \rho_p e^{i\mu_p} = \rho_p \cos \mu_p + i \sin \mu_p \\ &= \int_0^{2\pi} \cos p(\theta - \mu) f(\theta) d\theta + i \int_0^{2\pi} \sin p(\theta - \mu) f(\theta) d\theta \end{aligned} \quad (2.26)$$

$$\mu_p = \alpha_p + i\beta_p \quad (2.27)$$

When $p = 1$, we get $\alpha_1 = \rho$ and $\beta_1 = 0$, since $\int_0^{2\pi} \cos \theta d\theta = \rho \cos \mu$, $\int_0^{2\pi} \sin \theta d\theta = \rho \sin \mu$, so that $\int_0^{2\pi} \cos(\theta - \mu) d\theta = \rho$ and $\int_0^{2\pi} \sin(\theta - \mu) d\theta = 0$.

Considering equations (2.23)-(2.27) and the characteristic function definitions (2.19)-(2.20) it is easy to note that ϕ_p , α_p and β_p are the population versions of the p^{th} sample trigonometric moments m'_p , \bar{C}_p and \bar{S}_p defined by (2.11) and (2.12). Hence the characteristic function can be identified by the p^{th} trigonometric moments and

vice versa.

Replacing \bar{R} with ρ in the sample quantities defined in the previous Section, we can easily obtain the population measures reported below. In particular, the *circular variance* of Θ is defined by

$$v = 1 - \rho, \quad 0 \leq v \leq 1; \quad (2.28)$$

the *circular standard deviation* is defined, *not* as \sqrt{v} but as

$$\sigma = [-2 \log(1 - v)]^{\frac{1}{2}} \equiv (-2 \log \rho)^{\frac{1}{2}}$$

.

Finally, the population measures of skewness and kurtosis are defined, respectively, as

$$s = \beta_2 / (1 - \rho)^{3/2}$$

and

$$K = (\alpha_2 - \rho^4) / (1 - \rho)^2.$$

2.4 Statistical Approaches to Model Circular Data

There are many approaches to analyze circular data. The main ones are the *Embedding* approach, the *Intrinsic* approach and the *Wrapping* approach.

2.4.1 The embedding approach

In the embedding approach the sample space is considered as a part of a larger space and the distributions on the \mathbb{S}^{p-1} (the circular sample space) can be obtained by radial projection of the in line distributions on \mathbb{R}^p .

Projected Normal Distribution

Let \mathbf{Y} be a random vector in \mathbb{R}^p such that $Pr(\mathbf{Y} = 0) = 0$, then $\|\mathbf{Y}\|^{-1}\mathbf{Y}$ (with $\|\mathbf{Y}\|$ is the norm of \mathbf{Y}) is the corresponding *projected distribution* on \mathbb{S}^{p-1} . For instance, if \mathbf{Y} has a p -variate Normal distribution with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$, $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the corresponding projected distribution on \mathbb{S}^{p-1} is called *projected Normal distribution*, $PN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

The most used distribution in the embedding approach is surely the projected normal distribution in case $p = 2$, $PN_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, that is obtained projecting the bivariate in

line Normal distribution $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

The *probability density function* of the projected Normal distribution $PN_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ (Mardia, 1972) is:

$$Pr(\theta; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{\psi(\theta; \mathbf{0}, \boldsymbol{\Sigma}) + |\boldsymbol{\Sigma}|^{1/2} D(\theta) \Psi(D(\theta)) \psi(|\boldsymbol{\Sigma}|^{-1/2} (\mathbf{Y}^T \boldsymbol{\Sigma}^{-1} \mathbf{Y})^{-1/2} \boldsymbol{\mu} \wedge \mathbf{Y})}{\mathbf{Y}^T \boldsymbol{\Sigma}^{-1} \mathbf{Y}} \quad (2.29)$$

where $\psi(\cdot; \mathbf{0}, \boldsymbol{\Sigma})$ denotes the probability density function of $N_2(\mathbf{0}, \boldsymbol{\Sigma})$, ψ and Ψ denote, respectively, the probability density function and the cumulative density function of $N(0, 1)$, $\mathbf{Y} = (\cos \theta, \sin \theta)^T$, $\boldsymbol{\mu} \wedge \mathbf{Y} = \mu_1 \sin \theta - \mu_2 \cos \theta$ with $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$ and

$$D(\theta) = \frac{\mathbf{Y}^T \boldsymbol{\Sigma}^{-1} \mathbf{Y}}{(\mathbf{Y}^T \boldsymbol{\Sigma}^{-1} \mathbf{Y})^{1/2}}$$

The principal *properties* of the projected Normal distribution are:

- (i) the distribution $PN_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ reduces to the uniform distribution if and only if $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_2$, with \mathbf{I}_2 is the (2×2) identity matrix.
- (ii) projected Normal distribution can be bimodal and/or asymmetrical.

In general, the densities obtained by embedding a generic distribution on \mathbb{R}^p onto \mathbb{S}^{p-1} , can turn out to be very complicated and hence obtaining the likelihood-based inference can be extremely challenging. Moreover, most of the literature is focused on developing statistical methods for the projected Normal distribution only, which is a significant limitation of the embedding approach.

2.4.2 The intrinsic approach

In the intrinsic approach, the circle is used as the sample space. The directions are represented as points on the circle and probability distributions are defined on the circle directly. The main probability distributions obtained from this approach are the Uniform, Cardioid and von Mises distributions. In the following we present, for each of them, the probability density function, the distribution function and their main properties.

The circular Uniform distribution U_c

The most basic distribution on the circle is the *Uniform distribution*. It is the unique distribution on the circle which is invariant under rotation and reflection. With this model, all directions between 0° and 360° are equally likely.

The uniform model for circular data becomes relevant as a *null model*, against which several alternative (unimodal, multimodal) models can be tested.

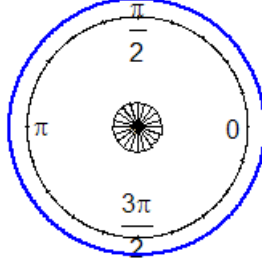


Figure 2.2: Density plot of circular Uniform distribution

The *Probability density function* is given by

$$f(\theta) = \frac{1}{2\pi}, \quad 0 \leq \theta \leq 2\pi, \quad (2.30)$$

while the *Distribution function* is

$$F(\theta) = \frac{\theta}{2\pi}, \quad 0 \leq \theta \leq 2\pi. \quad (2.31)$$

Thus, for $\alpha \leq \beta \leq \alpha + 2\pi$ it holds that

$$Pr(\alpha < \theta \leq \beta) = \frac{\beta - \alpha}{2\pi} \quad (2.32)$$

i.e. the probability is proportional to the arc length. Its characteristic function is

$$\phi_p = \begin{cases} 1, & p = 0 \\ 0, & p \neq 0 \end{cases} \quad (2.33)$$

Follow from equations (2.19), (2.23) and (2.28), that $\rho = 0$, so $v = 1$ and there is no concentration about any particular direction. The density plot is reported in Figure 2.2.

The *Moments* and some *properties* are given below:

- (i) the mean direction μ is undefined;
- (ii) the mean resultant length is $\rho = 0$;
- (iii) the circular dispersion is $\delta = \infty$;
- (iv) under a mild condition (see Section 4.3.1 Mardia (1972)), for any independent and identically distributed random variables $\Theta_1, \dots, \Theta_n$, $S_n = \Theta_1 + \dots + \Theta_n$ tends to the uniform distribution as $n \rightarrow \infty$. Furthermore, let Θ_1 be distributed uniformly and let Θ_2 have any distribution whatsoever; if Θ_1 and Θ_2 are independently distributed, then the characteristic function of $(\Theta_1 + \Theta_2)$ is given by (2.33). Hence, by the characteristic function uniqueness property, $(\Theta_1 + \Theta_2)$ is distributed uniformly.

The Cardioid Distribution $C(\mu, \rho)$

The cardioid distribution was introduced by Jeffreys (1948). This is a symmetric unimodal two-parameter distribution, sometimes referred to as the Cosine distribution.

The *probability density function* is given by

$$f(\theta) = \frac{1}{2\pi} \{1 + 2\rho \cos(\theta - \mu)\}, \quad 0 \leq \theta < 2\pi, \quad 0 \leq \rho \leq 1/2 \quad (2.34)$$

while the *distribution function* is

$$F(\theta) = (\rho/\pi) \sin(\theta - \mu) + \theta/(2\pi), \quad 0 \leq \theta \leq 2\pi \quad (2.35)$$

The *Moments* and some *properties* are:

- (i) the mean direction is μ ;
- (ii) the mean resultant length is $\rho(\leq 1/2)$;
- (iii) the circular dispersion is $\delta = 1/(2\rho^2)$;
- (iv) as $\rho \rightarrow 0$, the distribution converges to the uniform distribution U_c .

The von Mises Distribution $VM(\mu, h)$

This distribution was introduced by von Mises (1918) in order to study the deviations of measured atomic weights from integral values. From the statistical inference point of view, the von Mises is one of the most famous and used distributions on the circle. This is a symmetric (about $\theta = \mu$) unimodal distribution which is commonly used to modeling unimodal samples of circular data.

The *probability density function* - $VM(\mu, h)$ is

$$f(\theta) = \frac{1}{2\pi J_0(h)} e^{h \cos(\theta - \mu)} \quad (2.36)$$

where

$$J_0(h) = \frac{1}{2\pi} \int_0^{2\pi} e^{h \cos \theta} d\theta \quad (2.37)$$

is the modified Bessel function of the first kind and order zero. The function J_0 has power series expansion given by

$$J_0(h) = \sum_{r=0}^{\infty} \frac{1}{(r!)^2} (h/2)^{2r} \quad (2.38)$$

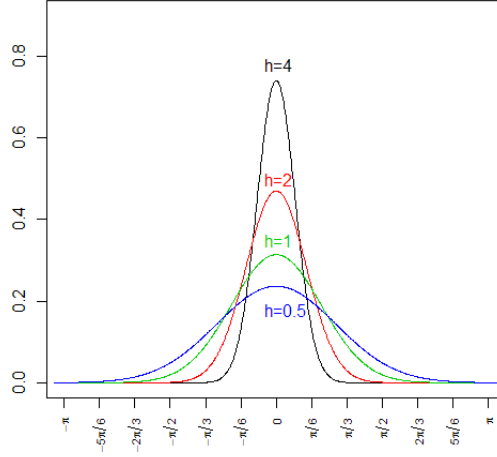


Figure 2.3: von Mises distribution with concentration parameter equal to 0.5, 1, 2, 4.

The parameter μ is the *mean direction* and the parameter h is known as the *concentration parameter*. A graphical representation of von Mises with different concentration parameter is reported in Figure 2.3. The mean resultant length ρ is equal to $A_1(h)$ where it is defined as $A_1(h) = J_1(h)/J_0(h)$ and, in general,

$$A_p(h) = J_p(h)/J_0(h), \quad p = 1, 2, \dots \quad (2.39)$$

with

$$J_p(h) = \sum_{r=0}^{\infty} \frac{1}{\Gamma(p+r+1)\Gamma(r+1)} (h/2)^{2r+p} \quad (2.40)$$

The *distribution function* is given by

$$F(\theta) = [2\pi J_0(h)]^{-1} \int_0^\theta e^{h \cos u} du \quad (2.41)$$

Main *moments* and *properties* are given below.

- (i) Mean direction: μ ;
- (ii) Mean resultant length: $\rho = A_1(h)$;
- (iii) Circular dispersion: $\delta = [hA_1(h)]^{-1}$;
- (iv) $\alpha_p = A_p(h)$ and $\beta_p = 0, p \geq 1$;
- (v) $\phi_p = e^{ip\mu \frac{J_p(h)}{J_0(h)}}$;
- (vi) as $h \rightarrow 0$, the distribution converges to the uniform distribution U_c ;
- (vii) as $h \rightarrow \infty$, the distribution tends to the point distribution concentrated in the direction μ ;

- (viii) mode is at $\theta = \mu$ and antinode (the opposite point in the circle at mode point) is at $\theta = \mu + \pi$. Note that $VM(\mu + \pi, h)$ and $VM(\mu, -h)$ are the same distribution, so to eliminate this indeterminacy of the parameters μ and h , it is usual to take $h \geq 0$;
- (xi) the ratio of the density at the mode to the density at the antinode is given by e^{2h} , so that the larger the value of h the greater is the clustering around the mode.

The von Mises distribution is perhaps the most prominent and used in literature, and its fame is due to the possibility of calculating the maximum likelihood parameter estimations. But one of the main drawbacks of this approach is that there are not many distributions available other than the von Mises and mixtures of von Mises distributions. Moreover, the extension to multivariate case for this distribution is still an open problem. Only in last recent work (Mardia *et al.*, 2007, 2008) we can see some applications of bivariate and trivariate von Mises distribution but their inference needs quite complex estimation procedure. These are the reasons why the wrapping approach, described in next section is often preferred to the previous ones.

2.4.3 Wrapping approach

The wrapping approach consists to wrap a known distribution in the real line around a circumference of a circle with a unit radius. The main characteristics of this approach is the flexibility: a rich class of distributions on the circle can be obtained using the wrapping technique because it is possible to wrap any known distribution in the real line onto the circle.

Wrapped distribution -general

Let Y be a in line random variable with probability density function $f(y)$, the corresponding circular random variable X_w is obtained by the following wrapping procedure:

$$X_w = Y(mod 2\pi) \quad (2.42)$$

The *probability density function* $f_w(\theta)$ of X_w is obtained by wrapping $f(y)$, defined on \mathbb{R} , around the circumference of a circle of unit radius, that is

$$f_w(\theta) = \sum_{k=-\infty}^{\infty} f(\theta + 2k\pi), \quad 0 \leq \theta < 2\pi \quad (2.43)$$

with corresponding *distribution function* given by

$$F_w(\theta) = \sum_{k=-\infty}^{\infty} F(\theta + 2k\pi) - F(2k\pi), \quad 0 \leq \theta < 2\pi \quad (2.44)$$

where F is the distribution function of Y .

The main *properties* are:

- (a) $(y_1 + y_2)_w = y_{1w} + y_{2w}$; i.e. the wrapping procedure is a homomorphism from \mathbb{R} to the circle.
- (b) If the characteristic function of Y is ϕ , then the characteristic function $\{\phi_p : p = 0, \pm 1, \dots\}$ of X_w is given by

$$\phi_p = \phi(p). \quad (2.45)$$

To see this:

$$\phi_p = \int_0^{2\pi} e^{ip\theta} f_w(\theta) d\theta = \sum_{k=-\infty}^{\infty} \int_{2\pi k}^{2\pi(k+1)} e^{ip\theta} f(\theta) d\theta = \int_{-\infty}^{\infty} e^{ipy} f(y) dy = \phi(p).$$

- (c) If ϕ is integrable then Y has a density and

$$f_w(\theta) = \sum_{k=-\infty}^{\infty} f(\theta + 2k\pi) = \frac{1}{2\pi} \left\{ 1 + 2 \sum_{p=1}^{\infty} (\alpha_p \cos p\theta + \beta_p \sin p\theta) \right\} \quad (2.46)$$

where $\phi(p) = \alpha_p + i\beta_p$. This result follows from (2.43) and from (2.44) and from considering the series $\sum_{p=1}^{\infty} (\alpha_p^2 + \beta_p^2)$ convergent because it holds that

$$\sum_p |\phi_p|^2 \leq \sum_p |\phi_p| \leq \int_{-\infty}^{\infty} |\phi(p)| dp.$$

- (d) If Y is infinitely divisible, then X_w is infinitely divisible.
- (e) There are (infinitely) many distributions on the line which can be wrapped onto any given distribution on the circle. To see this, let g be the probability density function of a distribution on the circle and define a probability density function on the line by

$$f(y) = p_r g(y), \quad 2r\pi < y \leq 2\pi(r+1), \quad r = 0, \pm 1, \pm 2, \dots$$

where p_r are any non-negative numbers such that $\sum_{r=-\infty}^{\infty} p_r = 1$. Then $f_w = g$.

Now, in the following subsections, we consider the main wrapped distributions.

Wrapped Poisson Distribution

Consider an in line random variable on the integers Y . Using a modified wrapping procedure, we can obtain the discrete wrapped random variable as:

$$X_w = 2\pi Y (\text{mod } 2\pi m) \quad (2.47)$$

with support on the lattice $\{2\pi r/m : r = 0, 1, \dots, m-1\}$ on the circle. In other words, just as the reduction modulo 2π wraps the in line onto the circle, so (if m is

a positive integer) the reduction modulo $2\pi m$ wraps the integer onto the group of m^{th} roots of 1, regarded as a subgroup of the circle (Mardia and Jupp, 1999).

The *probability density function* of X_w is given by:

$$Pr\left(x_w = \frac{2\pi r}{m}\right) = \sum_{k=-\infty}^{\infty} p(r + km), \quad r = 0, 1, \dots, m-1 \quad (2.48)$$

where p is the probability function of the in line variable Y .

In particular, if Y has the Poisson distribution with mean λ , then, from (2.48), X_w has the *wrapped Poisson distribution* with probability function given by:

$$Pr\left(x_w = \frac{2\pi r}{m}\right) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{r+km}}{(r + km)!}, \quad r = 0, 1, \dots, m-1 \quad (2.49)$$

From (2.45), the characteristic function of θ is

$$\phi_p = \exp\{\lambda(1 - e^{2\pi ip/m})\}$$

.

Wrapped Cauchy Distribution

Let Y be the Cauchy in line distribution with probability density function

$$f(y; \mu, a) = \frac{1}{\pi} \frac{a}{a^2 + (y - \mu)^2}, \quad -\infty < \mu < \infty, \quad a > 0;$$

and characteristic function is $e^{-a|p| - ip\mu}$. From (2.46) we are able to derive the *wrapped Cauchy distribution*, $WC(\mu, \rho)$, with density

$$f_w(\theta; \mu, \rho) = \sum_{k=-\infty}^{\infty} f(\theta + 2\pi k; \mu, a) = \frac{1}{2\pi} \left\{ 1 + 2 \sum_{p=1}^{\infty} \rho^p \cos p(\theta - \mu) \right\} \quad (2.50)$$

where $\rho = e^{-a}$.

Properties:

- (i) the mean direction is $\mu(\text{mod } 2\pi)$ and the mean resultant length is ρ ;
- (ii) the $WC(\mu, \rho)$ distribution is unimodal and symmetric about μ ;
- (iii) as $\rho \rightarrow 0$ it tends to uniform distribution and as $\rho \rightarrow 1$ it becomes concentrated at the point μ .

Wrapped Normal distribution

This is a symmetric unimodal two-parameter distribution which can be obtained by wrapping the in line Normal distribution $N(\tilde{\mu}, \sigma^2)$ around the circle, with

$$\sigma^2 = -2 \log \rho$$

i.e.

$$\rho = e^{-\sigma^2/2} \quad (2.51)$$

From the (2.43), the probability density function of the *wrapped Normal distribution*, $WN(\mu, \rho)$, is

$$f_w(\theta) = \frac{1}{\sigma\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} \exp\left\{-\frac{(\theta - \mu + 2k\pi)^2}{2\sigma^2}\right\} \quad (2.52)$$

Since the characteristic function of $N(\tilde{\mu}, \sigma^2)$ is given by $\phi(p) = e^{(i\mu p - p^2\sigma^2/2)}$, from the property (b) of the wrapped distribution derives that

$$\phi_p = e^{i\mu p - p^2\sigma^2/2}, \quad \alpha_p = e^{-p^2\sigma^2/2} \cos p\mu, \quad \beta_p = e^{-p^2\sigma^2/2} \sin p\mu$$

Using the previous quantities in equation (2.46), we obtain an useful representation of the density function (2.52) as

$$f_w(\theta) = \frac{1}{2\pi} \left\{ 1 + 2 \sum_{p=1}^{\infty} \rho^{p^2} \cos p(\theta - \mu) \right\} \quad (2.53)$$

For practical purposes, the probability density function (2.53) can be approximated adequately by the first three terms when $\sigma^2 > 2\pi$, while for $\sigma^2 \leq 2\pi$ the term with $k = 0$ of (2.52) gives a reasonable approximation, (see Mardia and Jupp, 1999, p.50).

Moments and properties

- (i) the mean direction is $\mu = \tilde{\mu}(\text{mod } 2\pi)$;
- (ii) the mean resultant length is $\rho = e^{-\sigma^2/2}$. Properties (i) and (ii) indicate a correspondence between parameters of in line and circular distribution.
- (iii) Circular dispersion: $\delta = (1 - \rho^4)/(2\rho^2)$;
- (iv) as $\rho \rightarrow 0$, the distribution converges to the uniform distribution U_c ;
- (v) as $\rho \rightarrow 1$, the distribution tends to the point distribution concentrated in the direction μ ;
- (vi) the mode is at $\theta = \mu$;
- (vii) if $\theta \sim WN(\mu, \rho)$ then $(\theta - \psi) \sim WN(\mu - \psi, \rho)$;

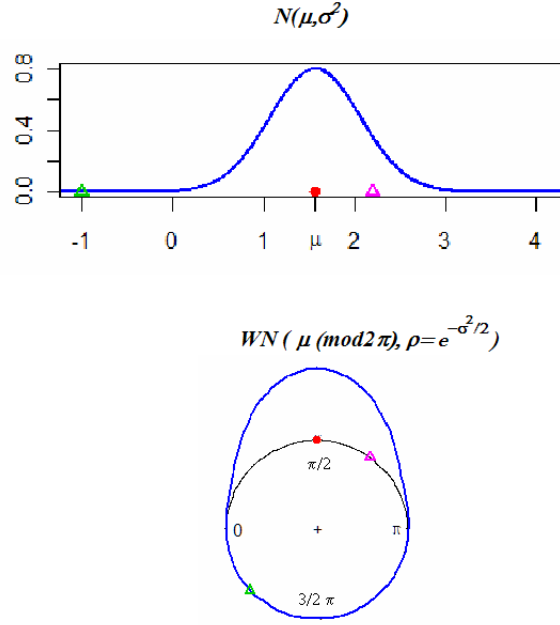


Figure 2.4: In line Normal distribution (top panel) and corresponding wrapped Normal distribution (bottom panel).

- (viii) from property (a) of the wrapped distributions and from equation (2.51) it follows that if Θ_1 and Θ_2 are two independent variables distributed as wrapped Normal, $\Theta_i \sim WN(\mu_i, \rho_i)$ $i = 1, 2$, then

$$(\Theta_1 + \Theta_2) \sim WN(\mu_1 + \mu_2, \rho_1 \rho_2).$$

Here we present only some of the circular distributions that we can obtain using wrapping method but in theory it is possible to derive a large (infinite) number of circular distributions, as many as the in line distributions are. This is of course an important advantage respect to the intrinsic approach. Moreover, in the majority of the wrapped distribution there is a correspondence between the in line parameters and corresponding wrapped parameters. This allows a easy interpretability of circular parameters and, then, of the all circular inference results.

2.4.4 Relationship among approaches and circular distributions

Considering the von Mises distribution, $VM(\mu, h)$, when $h = 0$ it correspond to the circular Uniform distribution. Moreover, the approximation $e^y \simeq 1 + y$ shows that for small h , the von Mises can be approximate to Cardioid distribution:

$$VM(\mu, h) \simeq C(\mu, h/2) \quad (2.54)$$

Thus a von Mises distribution with a small concentration parameter can be approximated by the Cardioid distribution with the same mean direction and mean resultant length equal to $h/2$. For large h , instead, it can be shown that, given $\Theta \sim VM(\mu, h)$, we have

$$[h^{-1/2}(\Theta - \mu)] \sim N(0, 1), h \rightarrow \infty \quad (2.55)$$

More generally, any von Mises distribution can be approximated by a wrapped Normal distribution:

$$VM(\mu, h) \cong WN(\mu, A_1(h)), \quad h \rightarrow \infty \quad (2.56)$$

where the $A_1(h)$ is defined by equation (2.39). Although the approximation (2.56) was derived as a first-order approximation for large h , Kent (1978) has shown that the approximation holds to a higher order in h ; more precisely:

$$f_{VM}(\theta; \mu, h) - f_{WN}(\theta; \mu, A_1(h)) = O(h^{-1/2}), \quad h \rightarrow \infty \quad (2.57)$$

where f_{VM} and f_{WN} are the densities of the von Mises $VM(\mu, h)$ and the wrapped Normal $WN(\mu, A_1(h))$ distribution, respectively. Stephens (1963) has numerically verified that the approximation (2.57) is satisfied for intermediate values of h . The worst match between a von Mises distribution and the wrapped Normal distribution occurs for $h \cong 1.4$, but also in this case the two distributions are very close. The von Mises is also close to the wrapped Cauchy distribution $WC(\mu, A_1(h))$ with the same mean direction and mean resultant length. This means that statistician can choose the most appropriate approach and, consequently, the most appropriate distributions according to his purpose. In particular, from statistical inference point of view, the intrinsic approach with von Mises distribution is the most suitable method to obtain the maximum likelihood parameter estimates. With the other approaches, instead, this result is unfeasible. On the other hand, for practical purposes, where the flexibility and interpretability are often the needful requirements, the wrapping approach performs better. In the last years, in fact, several developments for the wrapping approach have been focused on both to extend the wrapped family of circular distributions and to improve their inferential results.

2.5 Discussion

In this chapter we illustrate some basic tool to handle circular data. We first give measurements of location, concentration and the other descriptive statistics, then, we provide some element of circular distribution theory. In the second part of the chapter, we describe the three different approaches to model circular data. In particular, for each of them, we illustrate the main circular distributions providing descriptive and probabilistic characteristics and properties. Finally, the relationships between the approaches and related distributions are discussed. In particular, from a statistical inference point of view, the intrinsic approach with von Mises distribution

is the only one that allows a feasible maximum likelihood estimate of the parameters. Conversely, the wrapping approach, with its many circular distribution, appears particularly suitable for the practical purpose considering the ease of interpreting of the parameters. Finally, the embedding approach with its projected distributions is the method that arises difficulties in terms of both inference results and parameters interpretation.

In the next chapter we deeply analyze the wrapping approach in order to illustrate the main features and properties of this method. Our study aims to provide a method for extending to circular data the inferential procedures and models applied to in line data. In particular, we aim to apply to circular data, procedures and models of spatial and spatiotemporal contexts.

Chapter 3

Wrapping Approach

In the previous chapter we anticipate as the wrapping approach builds models for circular data by wrapping an in line density function $f(y)$, defined on \mathbb{R} , around the circle obtaining the wrapped density function $f_w(x)$ defined by

$$f_w(x) = \sum_{k=-\infty}^{\infty} f(x + 2k\pi) \quad 0 \leq x < 2\pi. \quad (3.1)$$

Thus, given a circular random variable X defined in $[0, 2\pi)$, through the transformation $(X + 2K\pi)$, with unobservable variable $K \in \mathbb{Z}$, we extend the support of X to \mathbb{R} so that we can apply an *in line* density function $f(y)$ to the argument $(X + 2K\pi)$. In other words, X represents a wrapped version of the *in line* random variable Y , which has probability density function $f(y \mid \Psi)$ that depends on an unknown parameter vector Ψ .

The great advantage of the class of *wrapped models*, and in particular of the wrapped Normal model, is the possibility of extending to circular multivariate variables and processes in a easy way. For example, suppose \mathbf{Y} is a d -dimensional random vector following the multivariate Normal distribution with probability density function $f(y \mid \mu, \Sigma)$. The wrapping procedure $X_w = Y(\text{mod} 2\pi)$ can be applied *component-wise* to obtain a d -dimensional multivariate distribution on the circle, with density

$$f_w(x_1, \dots, x_d) = \sum_{k_1=-\infty}^{\infty} \cdots \sum_{k_d=-\infty}^{\infty} f(x_1 + 2k_1\pi, \dots, x_d + 2k_d\pi) \quad (3.2)$$

Marginally, each variable X_i has a univariate wrapped Normal distribution, while the correlation structure is defined by the variance-covariance matrix, Σ of the in line distribution $f(y)$.

Suppose to have n replications of the d -dimensional Normal variable Y , in matrix form we can write:

$$\mathbf{Y} = \mathbf{X} + 2\mathbf{K}\pi \quad (3.3)$$

where $\mathbf{K} = \{K^{(1)}, \dots, K^{(n)}\}'$ and $\mathbf{X} = \{X^{(1)}, \dots, X^{(n)}\}'$ denotes a set of n d -dimensional circular random variables. Each $X^{(i)}$, $i = 1, \dots, n$ has density function on domain $[0, 2\pi)^d$ and such that $(X^{(i)}, X^{(j)}, i \neq j)$ are assumed to be pairwise independent and identically distributed .

More precisely, equation (3.3) can be rewritten as:

$$\begin{pmatrix} Y^{(1)} \\ \vdots \\ Y^{(i)} \\ \vdots \\ Y^{(n)} \end{pmatrix} = \begin{pmatrix} X^{(1)} + 2K^{(1)}\pi \\ \vdots \\ X^{(i)} + 2K^{(i)}\pi \\ \vdots \\ X^{(n)} + 2K^{(n)}\pi \end{pmatrix} = \begin{pmatrix} (X_1^{(1)} + 2K_1^{(1)}\pi, \dots, X_d^{(1)} + 2K_d^{(1)}\pi) \\ \vdots \\ (X_1^{(i)} + 2K_1^{(i)}\pi, \dots, X_d^{(i)} + 2K_d^{(i)}\pi) \\ \vdots \\ (X_1^{(n)} + 2K_1^{(n)}\pi, \dots, X_d^{(n)} + 2K_d^{(n)}\pi) \end{pmatrix} \quad (3.4)$$

Thus, the observed vector $X^{(i)}$ is one component of the vector pair $(X^{(i)}, K^{(i)})$, where the component $K^{(i)}$, referred to as the vector of wrapping coefficients, is unobservable.

This construction can be easily extended to circular stochastic processes, either in time or space. As a matter of fact, considering the d -components of (3.2) as variables of a process observed at d time points or d spatial locations and denoting by f_d the marginal joint density of this process, the corresponding wrapped density is given by:

$$f_d(x_1, \dots, x_d) = \sum_{k_1=-\infty}^{\infty} \cdots \sum_{k_d=-\infty}^{\infty} f(x_1 + 2k_1\pi, \dots, x_d + 2k_d\pi). \quad (3.5)$$

This class of models was studied in time series modeling by Breckling (1989), while general theory of wrapped circular distributions is introduced and discussed in Fisher (1993), Mardia and Jupp (1999) and Jammalamadaka and SenGupta (2001).

The *extension to multivariate* case given by (3.2) and (3.5) is one of the *major advantages* of this approach. Because of the growing amount of data, the possibility of modeling large data sets, often recorded in multivariate framework, is a focus of the modern methodologies. For this reason, in the last years, several attempts have been done to extend the most used circular distribution, the von Mises, to multivariate case. Only in the last recent literature we can find some results about this issue: Mardia *et al.* (2007), Hughes (2007) and Mardia *et al.* (2008), but even in these works, the modeled dimension is not higher than three and, however, with no few difficulties for the definition and estimation of the probability models.

Wrapping distributions and processes, therefore, provide a convenient and intuitive way to generate circular models. Moreover, the wrapped models have further advantages as the ease of interpretation due to the correspondence between the circular and wrapped parameters (see properties (i) and (ii) of the wrapped Normal distribution in Section 2.4.3). Besides, the in line distribution properties still hold under the wrapping procedure.

The difficulty in using such models stems from the complexity of inference: working directly with the density (3.1) or (3.2), the parameter estimation computing is intractable. A solution was developed by Fisher and Lee (1994) who used the Expectation-Maximization (EM) algorithm to obtain parameter estimates, treating the wrapping coefficients k 's as missing data. Though elegant this procedure presents high computational complexity: the E-step involves ratios of large infinite

sums which need to be approximated at each step making the algorithm computationally inefficient. More recently, Coles (1998) and Ravindran (2002) adopted a data augmentation approach to estimate the missing unobserved wrapping coefficients and the other parameters.

In this work we follow and extend the procedure proposed by Coles (1998) to a multivariate high dimensional process. In particular, we aim to deepen the knowledge about the wrapping approach and the role of the wrapping coefficients k 's in the parameter estimation in order to overcome the problems and difficulties in inference indicated, but not completely solved, in Coles (1998).

3.1 Parameter Estimation

The main difficulty in working with the wrapping approach is that the form of the density function is constituted by large sums, and cannot be simplified as close form. In order to solve this drawback, we assume that a circular distribution is obtained by wrapping onto the circle of an in line distribution. Therefore, if it is possible to *unwrap* the distribution on the circle and to obtain a distribution on the real line, it will be possible to use all the standard statistical techniques for variables defined on the real line. In other words, given a circular random variable X , it is always possible to write the corresponding *in line* random variable Y as $Y = (X + 2K\pi)$, where X is defined in $[0, 2\pi)$ and K is the unobservable variable representing the number of times Y is wrapped to obtain X . Therefore, if it is feasible to add the information on K , and thus *unwrap* X , then it is possible to work directly with Y .

In this sense, making inference on X , whose density function $f_w(y)$ is written in (3.1), is intractable. The in line density function $f(y)$, instead, is simpler and describes the density of the in line variable Y or, equivalently, of the vector pair (X, K) of which we observe only X . The equivalence

$$Y \equiv (X, K) \tag{3.6}$$

has to be interpreted as the joint distribution for (X, K) with argument $(X + 2K\pi)$ and parameters vector Ψ .

As K is a unobservable variable, a natural and convenient approach, that enables to overcome the uncertainty on K by writing conditional distribution, is the Bayesian via Markov Chain Monte Carlo (MCMC) methods. The basis of such methods is to simulate a Markov chain whose equilibrium distribution is proportional to a required function $\pi(t)$, which, in our case, is the posterior distribution of (K, Ψ) .

The most general MCMC method is the Metropolis-Hastings (M-H) algorithm that we briefly introduce.

Given a current value $t^{(i)} = t$, a proposed value t' is simulated from an arbitrary transition density $q(t' | t)$. Then, $t^{(i+1)}$ is adopted according to the following rule:

$$t^{(i+1)} = \begin{cases} t' & \text{with probability } \Delta(t^{(i)}, t') \\ t^{(i)} & \text{with probability } 1 - \Delta(t^{(i)}, t') \end{cases} \tag{3.7}$$

where

$$\Delta(t, t') = \min \left\{ 1, \frac{\pi(t')q(t | t')}{\pi(t)q(t' | t)} \right\} \quad (3.8)$$

In our situation, t is multivariate and defined by $t = \{\Psi, \mathbf{K}\}$, and the algorithm is applied in cycles to subsets of t . This is done simply replacing $\pi(t)$ in equation (3.8) with the appropriate conditional density, $\pi_m(t_m | t_{-m})$, and cycling through m , with a proposed transition density q_m for each component. For more details, see Smith and Roberts (1993). The implementation of the M-H algorithm is carried out by successively updating the components of Ψ , the parameter vector of the in line distribution, and the wrapping coefficient vector, \mathbf{K} , one component at a time. Note that, in the multivariate case, each vector $X^{(i)}$ carries its own vector $K^{(i)} = (k_1^{(i)}, \dots, k_d^{(i)})$ of missing components, so the number of unknown parameters in this model is large. But, recalling equation (3.4), the algorithm is simplified by the following identities:

$$Pr(\Psi | \mathbf{X}, \mathbf{K}) = \frac{Pr(\mathbf{X}, \mathbf{K} | \Psi)Pr(\Psi)}{Pr(\mathbf{X}, \mathbf{K})} \propto Pr(\mathbf{X}, \mathbf{K} | \Psi)Pr(\Psi) \quad (3.9)$$

and

$$\begin{aligned} Pr(k_j^{(i)} | \mathbf{X}, \mathbf{K}^{(-i)}, k_{-j}^{(i)}, \Psi) &= Pr(k_j^{(i)} | X^{(i)}, k_{-j}^{(i)}, \Psi) \\ &= \frac{Pr(X^{(i)}, k^{(i)} | \Psi)}{Pr(X^{(i)} | \Psi)} \\ &\propto Pr(X^{(i)}, k^{(i)} | \Psi), \quad i = 1, \dots, n; \quad j = 1, \dots, d \end{aligned} \quad (3.10)$$

Henceforth, we assume \mathbf{Y} and, then (\mathbf{X}, \mathbf{K}) , distributed as multivariate Normal densities with parameter vector $\Psi = \{\mu, \Sigma\}$, therefore, $Pr(X^{(i)}, k^{(i)} | \Psi)$ is a multivariate Normal density function, while $Pr(\mathbf{X}, \mathbf{K} | \Psi)$ is the product of such functions due to the conditional independence of the $X^{(i)}$. Substituting the posterior densities (3.9) and (3.10) into equation (3.8) leads to easy expressions for the acceptance probability of proposal transitions in the simulated Markov chain.

One possible choice for the transition density q_m is based on a random walk. In this case, for each component of Ψ a transition of the form

$$t' = t + \epsilon \quad (3.11)$$

is adopted where $\epsilon \sim Uniform[-a, a]$.

Instead, the transition model for each wrapping coefficient $k_j^{(i)}$ is a discrete random walk given by

$$k' = k + \epsilon \quad (3.12)$$

where $\epsilon = \{-1, 0, 1\}$ with probabilities $\{p, 1 - 2p, p\}$ respectively, for some opportune choice of p .

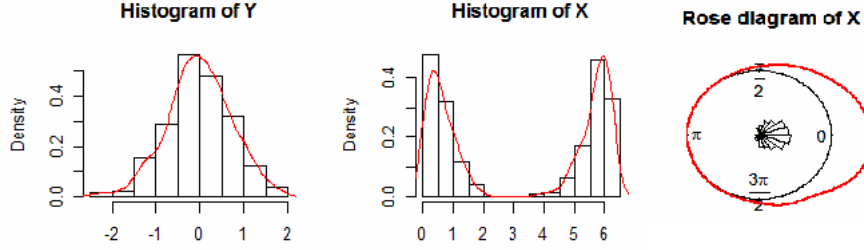


Figure 3.1: Univariate example 1: histogram of the in line data Y , simulated from $N(0, \pi/4)$, and histogram and rose diagram of the corresponding wrapped data.

Parameter	Univariate Example 1			Univariate Example 2		
	True	Posterior	95%	True	Posterior	95%
	Value	Median	Posterior Interval	Values	Median	Posterior Interval
μ	0	0.007	$[-0.016, 0.008]$	1	1.148	$[0.768, 1.425]$
σ^2	$\pi/4$	0.775	$[0.713, 0.822]$	π	2.578	$[2.367, 3.624]$

Table 3.1: Estimates of the two univariate wrapped Normal distributions plotted in Figure 3.1

For more details about estimation procedure see Coles (1998). Wide literature can be found for having a general overview about Bayesian inference by MCMC algorithms: among the most recent works we cite Gelman *et al.* (2004), Barbieri (1996), Gelman *et al.* (1996), Besag *et al.* (1995), Daniel and Gatsonis (1999), Gelfand *et al.* (1995), Gelfand and Smith (1990), Harville and Zimmerman (1996).

3.1.1 Simulated examples

In this section we present some simulated examples of univariate and multivariate cases in order to show the M-H algorithm for the parameter estimation procedure described in the previous section. In the univariate case we apply the M-H to two simulated data sets drawn from univariate Normal distributions, $N(0, \pi/4)$ and $N(1, \pi)$ respectively. These data sets are depicted in Figure 3.1-3.2.

For the first example we use quite flat priors for both parameters μ and σ^2 ; in particular a Normal prior $N(0, 10)$ for the mean parameter μ and an inverse Gamma prior $InvGamma(4, 10)$ for σ^2 . A different setting, instead, is adopted for the second example: a Normal prior $N(1, 0.5)$ for μ , and an $InvGamma(10, 40)$ for σ^2 that are more informative with respect to the priors of the first example. To decide this kind of setting we followed the suggestions written in Coles (1998), where the author recommends to use more informative priors when the data variability increases. The Markov chain Monte Carlo outputs for these two data sets are summarized in Table 3.1 and depicted in Figure 3.3.

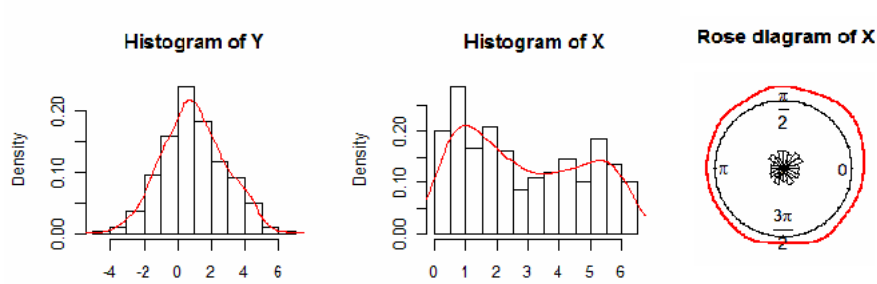


Figure 3.2: Univariate example 2: histogram of in line data Y, simulated from $N(1, \pi)$, and histogram and rose diagram of the corresponding wrapped data.

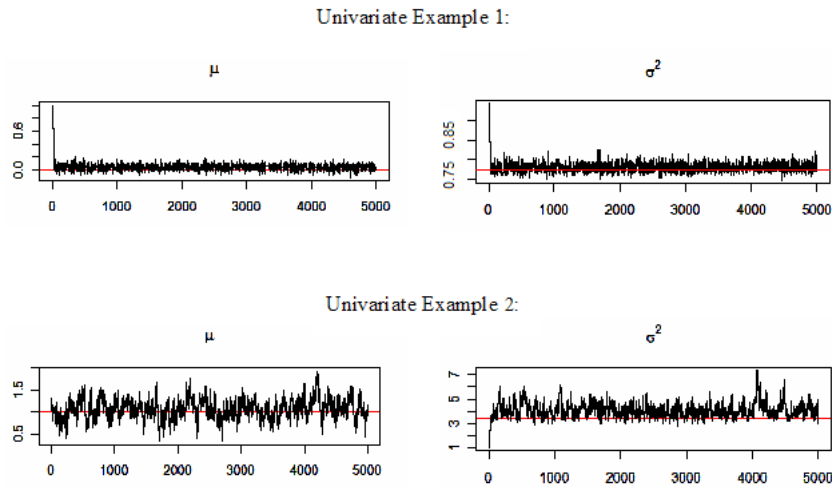


Figure 3.3: Traces of the estimated parameter chains of univariate examples 1 and 2

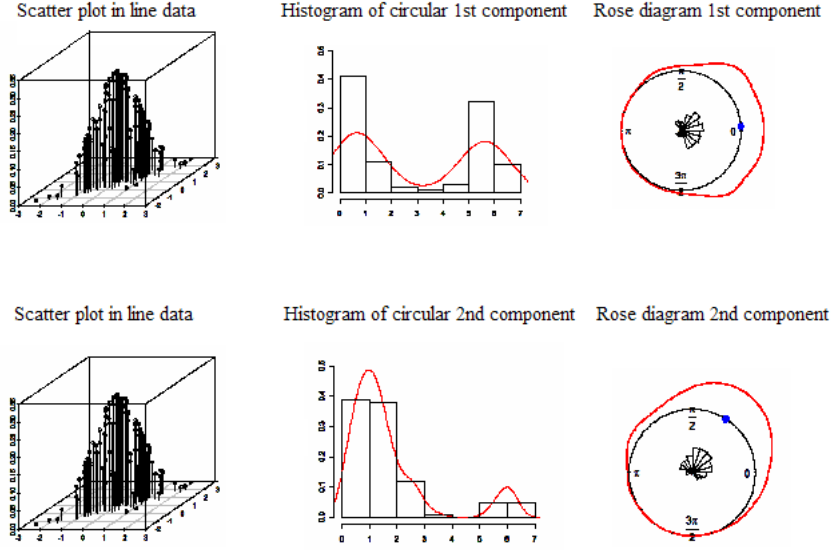


Figure 3.4: Bivariate example: scatter plot of the in line data Y , simulated from $N_2(\mu = (0, 1)', \Sigma)$, and histogram and rose diagram of the corresponding wrapped components.

Parameter	Bivariate Example		
	True Value	Posterior Median	95% Posterior Interval
μ_1	0	-0.092	[-0.167, 0.123]
μ_2	1	1.042	[0.986, 1.208]
σ_1^2	1	1.067	[0.967, 1.245]
σ_2^2	1	1.103	[0.920, 1.289]
σ_{12}	0.7	0.668	[0.562, 0.845]

Table 3.2: Parameter estimates of the bivariate wrapped Normal distribution plotted in Figure 3.4

More generally, we can generalize the parameter estimation procedure described in the previous section to the multivariate case by using an inverse Wishart prior for the variance-covariance matrix, Σ , of the multivariate wrapped Normal distribution. For this example, we simulated the in line data Y from a bivariate Normal distribution $N_2(\mu, \Sigma)$ and then applied the wrapping procedure componentwise in order to obtain the corresponding two circular components X_1 and X_2 . The simulated data specification are reported in Table 3.2 and some graphical representations are given in Figure 3.4.

For this bivariate case, the transition density for the variance-covariance matrix is an inverse Wishart with $d + 1$ degrees of freedom (where d is the dimension of the multivariate in line distribution) and scale matrix equal to the variance-covariance matrix at the previous iteration. More details about using the Wishart and inverse Wishart prior for variance-covariance matrix in MCMC algorithm can be found in

Parameter	Univariate Example 1	Univariate Example 2
	Z-score	Z-score
μ	0.169	1.983
σ^2	-0.234	-1.315

Table 3.3: Geweke Z-score diagnostic for the parameters of both univariate examples.

Gelman (2006) and Kass and Natarajan (2006); general details about the random walk transition density in Metropolis-Hastings algorithms can be read in Chib and Greenberg (1998) and in Ishwaran and Gatsonis (2000).

Table 3.2 shows that the parameter estimation produces good results also for the generalized bivariate Normal distribution, with posterior estimates very close to true values and small posterior intervals.

Through the examples just examined, we showed the easy applicability of the wrapping approach in both the univariate and the more general multivariate cases. Moreover, analyzing the results obtained from univariate examples, we remark good estimates in the first example where the simulated data have been drawn from a distribution with small variance (equal to $\pi/4$), whereas very poor estimates have been obtained from the second example data which has quite large variability (equal to π). In the first univariate example, in fact, the posterior interval lengths for μ and σ are respectively equal to 0.024 and 0.109; in the second example, instead, they are larger: 0.667 and 1.257, respectively, suggesting a low reliability of estimates. Besides, looking at the chain traces reported in Figure 3.3 (bottom plots), we realize the chain convergence failure. The autocorrelation function and the Geweke convergence diagnostic (Geweke, 1992), reported respectively in Figure 3.5 and in Table 3.3, confirm this failure. In the second example, in fact, the chains of both parameters have high correlation and the Z-score falls in the extreme tail of the standard Normal distribution, rejecting the null hypothesis of equality of means between the first and last part of the chain.

The estimate unreliability problem highlighted for data sets with large variability was very briefly discussed by Coles (1998). He asserts that when the in line data exhibit large variance, there is little information available in wrapped data to ascertain how many times an in line observation has been wrapped. In other words, when the data have large variance, determining the wrapping coefficients k 's is very difficult and all inferential results are instable.

In the next section, we analyze the central role of the k 's coefficients in the wrapping approach in order to detect when and how the identifiability problem for k 's arises. Moreover, we aim to use the acquired knowledge on k 's to improve inference and, consequently, the estimate reliability for any data sets.

To analyze the role and the behavior of the wrapping coefficients in inference and to verify the k 's identifiability problem for large variability data sets, we apply the parameter estimation procedure described in Section 3.1 to several simulated data sets. In particular, we first detect the data sets where the inference procedure

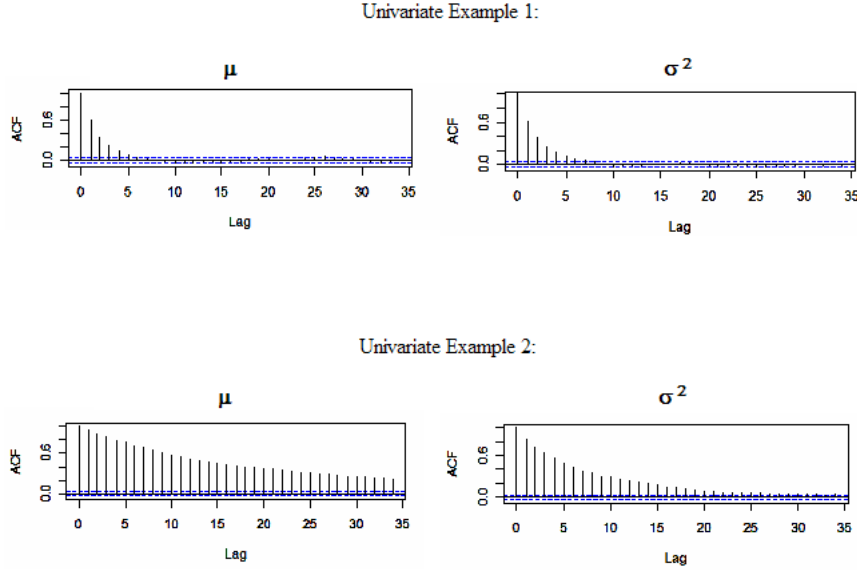


Figure 3.5: Autocorrelation functions of the estimated parameters of the univariate example 1 and 2.

fails, and, for these data, we investigate the role of the k coefficients in inference. For the first purpose, we take into account different data scenarios using data sets that differ both in dimension (number of replications) and in variability.

3.2 Sample Size Effect on Parameter Estimation

In this section we show some applications of the parameter estimation procedure in the case of univariate wrapped Normal model: in particular, here we investigate the effects of changing the sample size (or number of replications), n . Moreover, for all simulated data, we analyze the influence of the prior distributions on the posterior estimates.

We start our analysis considering four data sets that differ in the number of replications. As our aim here is to study exclusively the sample size effect on the estimates, we decide to use the same distribution of the univariate example 1 described in Section 3.1 that gave good results. For this purpose, we draw four samples from the in line Gaussian distributions with zero mean, variance equal to $\pi/4$ and sample size n equal to 10, 40, 100 and 300, respectively. Then we wrap these simulated data to obtain the corresponding wrapped Normal data sets.

In figure 3.6 we report the histograms of the in line variables and the rose diagrams with density estimates of the corresponding wrapped variables. We can note that when the number of replications is less than 40, the density shape is quite irregular, showing more than one mode, so in these cases, we expect a poor inference on parameter estimates. In cases (c) and (d), where the number of replications is instead more than 100, the data distribution is regular, and good estimation results are expected.

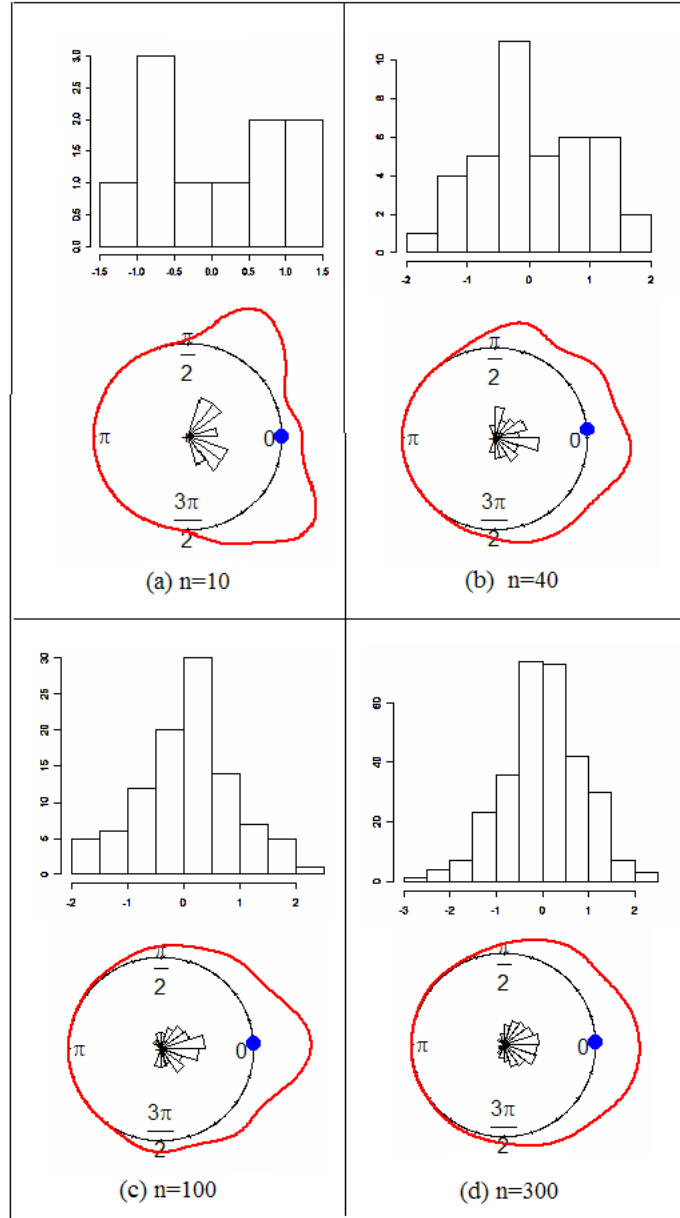


Figure 3.6: Histograms of the in line data simulated from $N(0, \pi/4)$ and corresponding rose diagrams, density estimates (red lines) and means (blue dots) of the wrapped data. The number of replications n is respectively equal to 10 (a), 40 (b), 100 (c) and 300 (d).

Parameter	Flat Priors				Informative Priors			
	Z-score (a)	Z-score (b)	Z-score (c)	Z-score (d)	Z-score (a)	Z-score (b)	Z-score (c)	Z-score (d)
μ	0.569	0.427	0.201	0.186	0.552	0.331	0.210	0.113
σ^2	-0.901	-0.723	-0.702	-0.526	-0.628	-0.567	-0.354	-0.298

Table 3.4: Geweke Z-score diagnostic (computed with burn-in of 1000 iteration) for the chains depicted in Figures 3.7 and 3.8

N	μ			σ^2		
	True Value	Posterior Median	95% Posterior Interval	True Values	Posterior Median	95% Posterior Interval
10	0	-0.012	[-0.511, 0.530]	$\pi/4=0.785$	0.626	[0.293, 1.438]
40	0	-0.077	[-0.356, 0.200]	$\pi/4=0.785$	0.786	[0.538, 1.217]
100	0	0.037	[-0.130, 0.208]	$\pi/4=0.785$	0.761	[0.569, 1.014]
300	0	-0.008	[-0.106, 0.092]	$\pi/4=0.785$	0.779	[0.671, 0.916]

Table 3.5: Posterior results for the four data sets plotted in Figure 3.6, obtained using the flat priors: $\mu \sim N(0, 25)$ and $\sigma^2 \sim InvGamma(4, 8)$

In tables 3.5 and 3.6 the results of inference for the four data set are reported. To obtain the estimates in Table 3.5 we use quite flat priors on μ and σ^2 : $\mu \sim N(0, 25)$ and $\sigma^2 \sim InvGamma(4, 8)$. Instead, to obtain the results in Table 3.6, we use more informative priors that are: $\mu \sim N(0, 2)$ and $\sigma^2 \sim InvGamma(4, 2)$. As expected, the smaller the n value is, the less accurate the estimates are. As a matter of fact, the 95% posterior interval lengths increase as the number of replications decreases. In the case of flat priors, for the parameter μ the length is 1.041 with $n = 10$ and 0.198 with $n = 300$; while in case of more informative priors we have a posterior interval length of 1.021 for $n = 10$ and of 0.196 for $n = 300$. We have basically the same behavior of the posterior interval length for the variance σ^2 .

Finally, viewing the chain plots in Figures 3.7-3.8, we note that the priors have no so much influence both on the parameter estimates, that mainly depend on the sample dimension n , and on the chains convergence, that occurs quite quickly. For both flat and more informative priors, in fact, the chains convergence occurs after only 1000 iteration and the Geweke diagnostic reported in Table 3.4 confirms this evidence.

N	μ			σ^2		
	True Values	Posterior Median	95% Posterior Interval	True Values	Posterior Median	95% Posterior Interval
10	0	-0.032	[-0.506, 0.515]	$\pi/4=0.785$	0.831	[0.278, 1.388]
40	0	-0.008	[-0.258, 0.268]	$\pi/4=0.785$	0.751	[0.534, 1.152]
100	0	-0.027	[-0.192, 0.138]	$\pi/4=0.785$	0.728	[0.569, 0.957]
300	0	-0.019	[-0.122, 0.074]	$\pi/4=0.785$	0.761	[0.655, 0.903]

Table 3.6: Posterior results for the four data sets plotted in Figure 3.6, obtained using the priors: $\mu \sim N(0, 2)$ and $\sigma^2 \sim InvGamma(4, 2)$

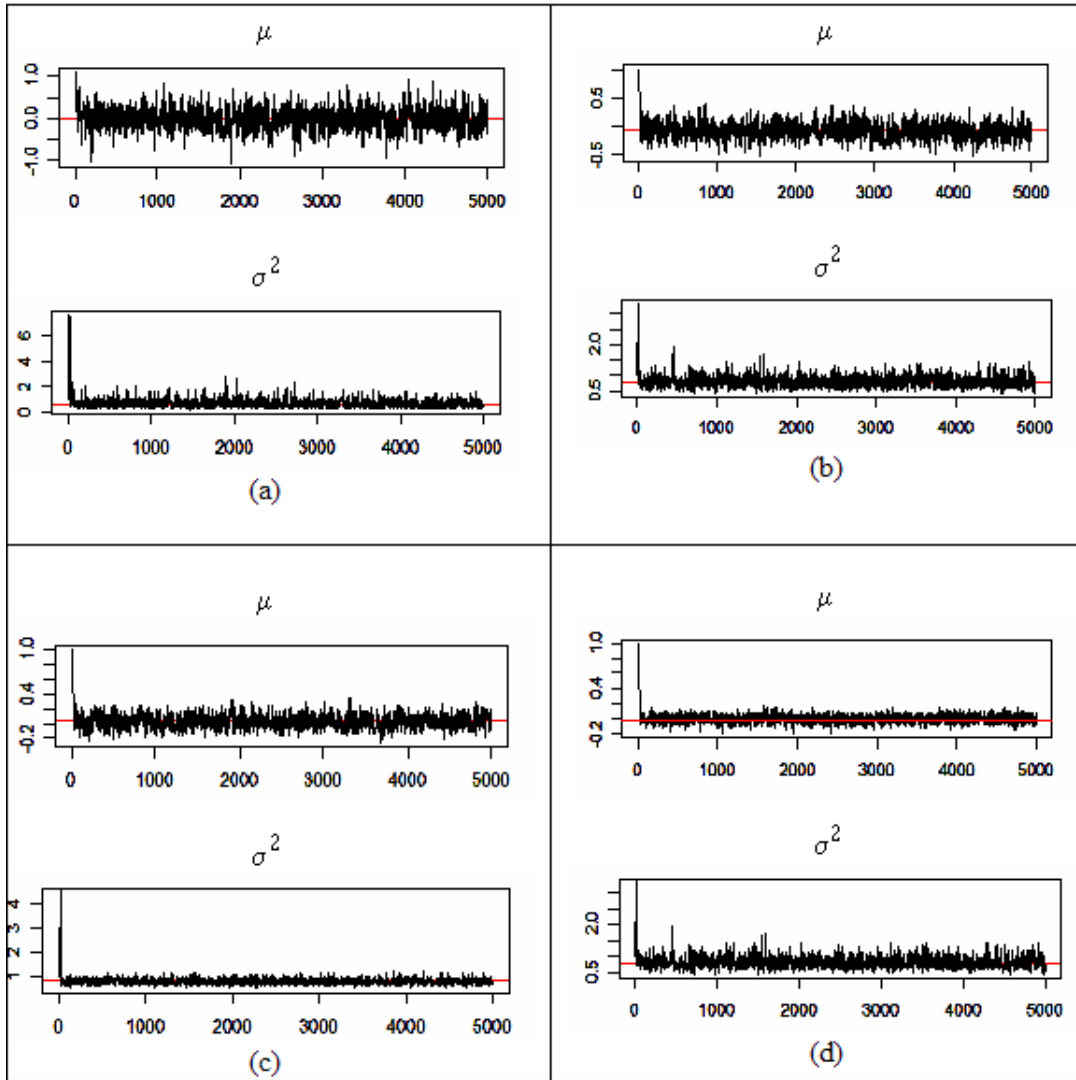


Figure 3.7: MCMC chain traces for μ and σ^2 parameters of the four simulated data sets. Outputs obtained with flat priors $\mu \sim N(0, 25)$ and $\sigma^2 \sim InvGamma(4, 8)$

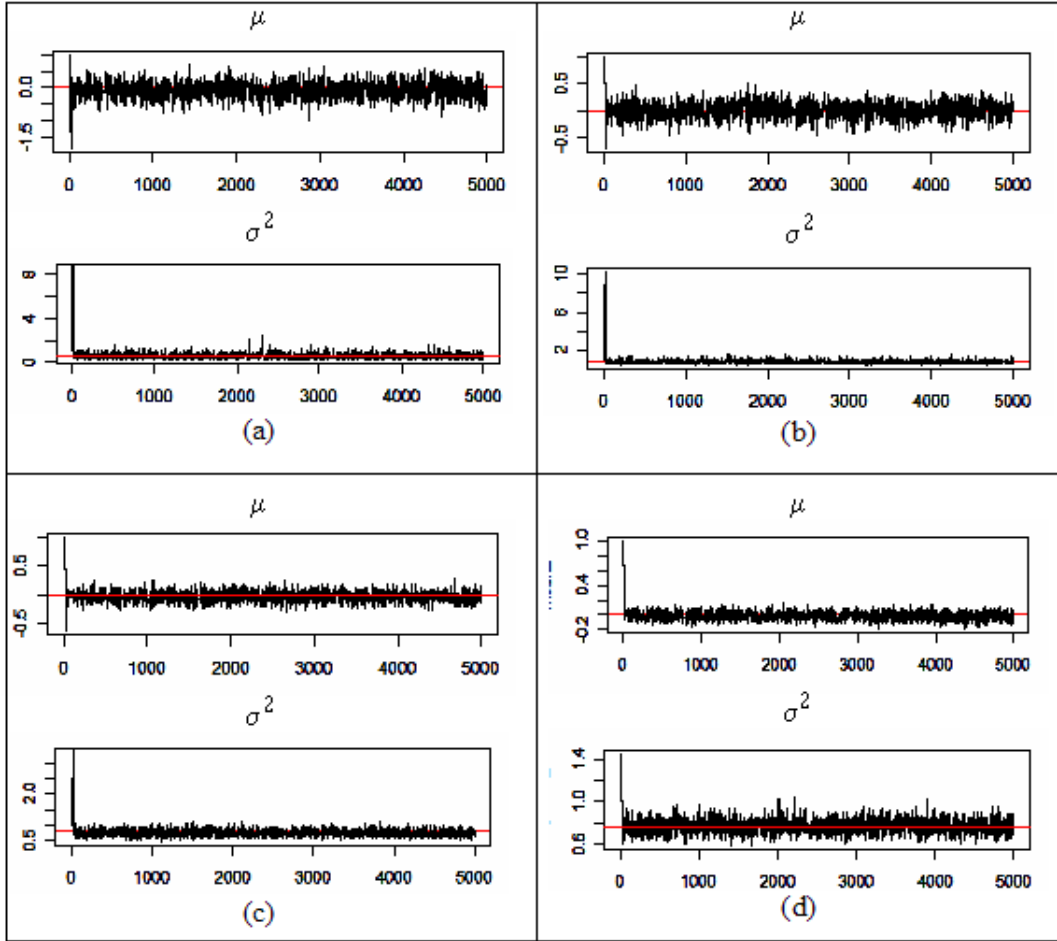


Figure 3.8: MCMC chain traces for μ and σ^2 parameters of the four simulated data sets. Outputs obtained with informative priors $\mu \sim N(0, 2)$ and $\sigma^2 \sim \text{InvGamma}(4, 2)$

σ^2	μ			σ^2		
	True Value	Posterior Median	95% Posterior Interval	True Values	Posterior Median	95% Posterior Interval
$\pi/10$	0	0.026	[-0.048, 0.151]	$\pi/10=0.314$	0.328	[0.283, 0.391]
$\pi/8$	0	0.002	[-0.052, 0.089]	$\pi/8=0.393$	0.372	[0.314, 0.457]
$\pi/6$	0	-0.071	[-0.172, 0.027]	$\pi/6=0.524$	0.495	[0.382, 0.656]
$\pi/4$	0	-0.003	[-0.122, 0.189]	$\pi/4=0.785$	0.837	[0.578, 1.033]
$\pi/2$	0	0.008	[-0.159, 0.182]	$\pi/2=1.571$	1.278	[0.856, 1.778]
π	0	-0.146	[-0.323, 0.345]	$\pi=3.142$	2.487	[2.179, 3.622]
$3\pi/2$	0	0.291	[-1.150, 1.537]	$3\pi/2=4.712$	5.956	[2.894, 9.115]

Table 3.7: Posterior results for the seven data sets plotted in Figure 3.7, obtained using the following flat priors: $\mu \sim N(0, 30)$ and $\sigma^2 \sim \text{InvGamma}(4, 8)$

3.3 Variability Effect on Parameter Estimation

Now we deal with the effect of modifying the data set variability on the parameter estimates. As in the previous examples, we consider several data sets that, in this case, differ in variability, while the number of replications n remains constant. For this purpose we choose the sample size n equal to 300, i.e. the sample size that in the previous subsection gave the best results. Besides, we carry out the variability effect analysis considering also the influence of the prior distributions on the posterior estimates.

We simulate seven wrapped Normal data sets by wrapping the in line data simulated from Gaussian Normal distributions that have zero mean and variance equal to $\pi/10$, $\pi/8$, $\pi/6$, $\pi/4$, $\pi/2$, π and $3\pi/2$, respectively.

In Figure 3.9 we can see the histograms and the rose diagrams of the seven data sets used for the analysis. It is worth noting plots (f) and (g): they clearly show that when the in line variance is larger than π the corresponding wrapped distribution is highly dispersed and seems very similar to the circular Uniform distribution.

The MCMC inference results for these seven data sets are summarized in Tables 3.7-3.8. In particular, Table 3.7 refers to the case of quite flat priors on the parameters, that is $\mu \sim N(0, 30)$ and $\sigma^2 \sim \text{InvGamma}(4, 8)$, while Table 3.8 considers the case of more informative priors that is $\mu \sim N(0, 1)$ and $\sigma^2 \sim \text{InvGamma}(4, 2)$.

At a first sight, we can say that the results in Table 3.8 are more accurate than those in Table 3.7. The estimates are closer to the true values and the posterior interval lengths are smaller when the priors are more informative. In particular, considering the parameter μ , the interval lengths are between 0.141 and 2.687 with flat priors and between 0.129 and 2.187 with more informative priors.

Finally, we report the parameter chain traces in Figures 3.10-3.11. The first figure refers to the case of flat priors, while the second one to the case of more informative priors. Comparing these figures and the Geweke convergence diagnostic results reported in Table 3.9, we realize that the priors have not so much influence on the chain convergence and on the estimates as the variability have: the Geweke

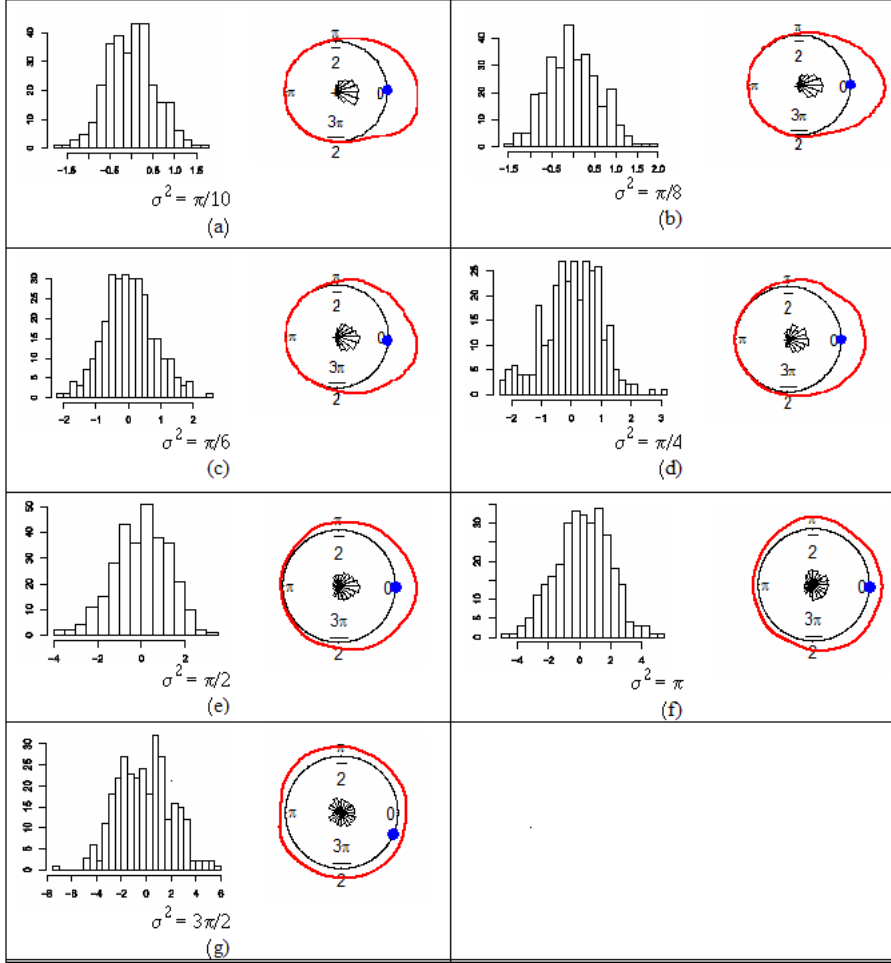


Figure 3.9: Histograms of the in line data simulated from $N(0, \sigma^2)$ and corresponding rose diagrams, density estimates (red lines) and means (blue dots) of wrapped Normal data. The variance σ^2 is respectively equal to $\pi/10$ (a), $\pi/8$ (b), $\pi/6$ (c), $\pi/4$ (d), $\pi/2$ (e), π (f) and $3\pi/2$ (g).

σ^2	μ			σ^2		
	True Value	Posterior Median	95% Posterior Interval	True Values	Posterior Median	95% Posterior Interval
$\pi/10$	0	0.037	[-0.028, 0.101]	$\pi/10=0.314$	0.325	[0.279, 0.382]
$\pi/8$	0	-0.006	[-0.062, 0.077]	$\pi/8=0.393$	0.379	[0.324, 0.447]
$\pi/6$	0	-0.079	[-0.161, 0.003]	$\pi/6=0.524$	0.505	[0.432, 0.596]
$\pi/4$	0	-0.003	[-0.101, 0.098]	$\pi/4=0.785$	0.787	[0.673, 0.933]
$\pi/2$	0	0.002	[-0.139, 0.152]	$\pi/2=1.571$	1.398	[1.193, 1.648]
π	0	-0.041	[-0.343, 0.290]	$\pi=3.142$	2.887	[2.379, 3.522]
$3\pi/2$	0	0.192	[-0.650, 1.537]	$3\pi/2=4.712$	5.012	[3.893, 8.436]

Table 3.8: Posterior results for the seven data sets plotted in Figure 3.7, obtained using the following informative priors: $\mu \sim N(0, 1)$ and $\sigma^2 \sim InvGamma(4, 2)$

Variance	Flat Priors		Informative Priors	
	Z-score	Z-score	Z-score	Z-score
	μ	σ^2	μ	σ^2
$\pi/10$	0.198	-0.154	0.167	-0.187
$\pi/8$	0.123	-0.267	0.231	-0.126
$\pi/6$	0.206	-0.312	0.210	-0.256
$\pi/4$	0.154	-0.212	0.189	-0.206
$\pi/2$	0.567	-0.498	0.379	-0.401
π	1.981	-1.785	1.657	-1.259
$3\pi/2$	3.526	-2.764	2.228	-2.029

Table 3.9: Geweke Z-score diagnostic (computed with burn-in of 5000 iterations) for the chains depicted in Figures 3.10 and 3.11

diagnostic fails and the posterior interval lengths increase when the variance is larger than $\pi/2$. Anyway, when the variance is not too large, for example in the cases of $\sigma^2 = \pi/2$ or $\sigma^2 = \pi$, the informative priors helps the MCMC algorithm to obtain more accurate estimates. To see this we can look at the posterior intervals related to these variances in Tables 3.7 and 3.8.

Summarizing the evidence that emerges from all these examples, we can state that the main problem for "goodness of inference" in the wrapping approach is due to the large variability of data. Besides, the introduction of more informative priors improves the estimates only when the variability is not too large, as between $\pi/2$ and π . Finally, from this simulation-based study, we have detected the data variability threshold, equal to $\pi/2$, over which the inference procedure of wrapping approach fails.

3.3 Variability Effect on Parameter Estimation

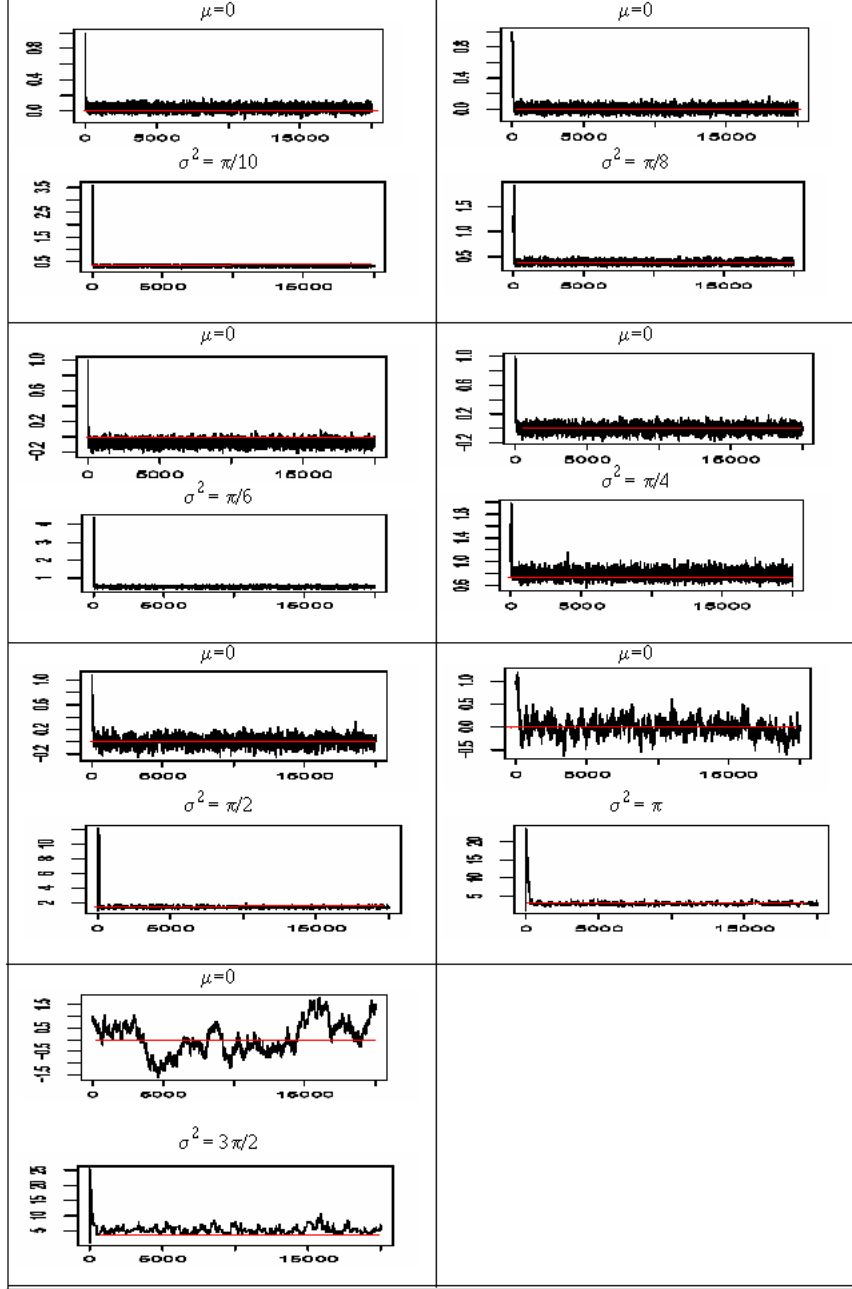


Figure 3.10: MCMC chain traces for μ and σ^2 parameters of the seven simulated data sets. Outputs obtained with flat priors $\mu \sim N(0, 30)$ and $\sigma^2 \sim \text{InvGamma}(4, 8)$

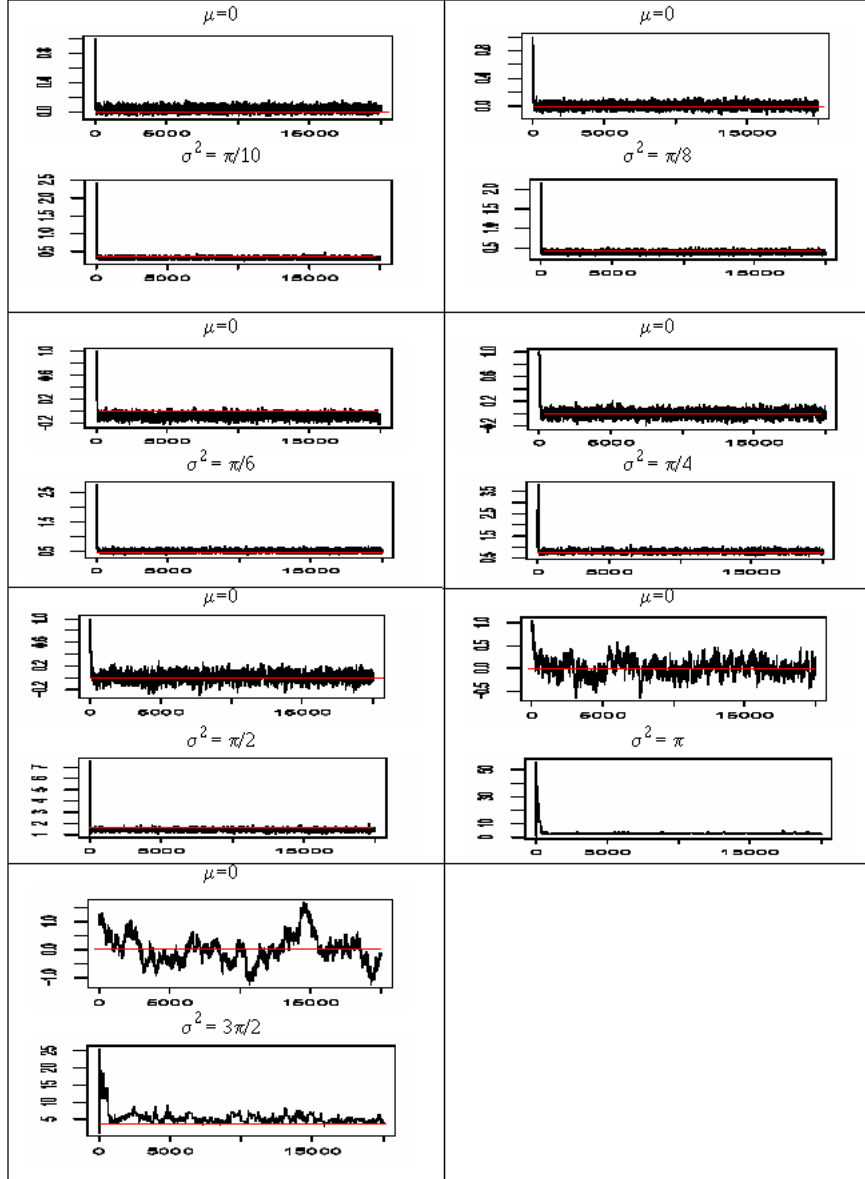


Figure 3.11: MCMC chain traces for μ and σ^2 parameters of the seven simulated data sets. Outputs obtained with informative priors: $\mu \sim N(0, 1)$ and $\sigma^2 \sim \text{InvGamma}(4, 2)$

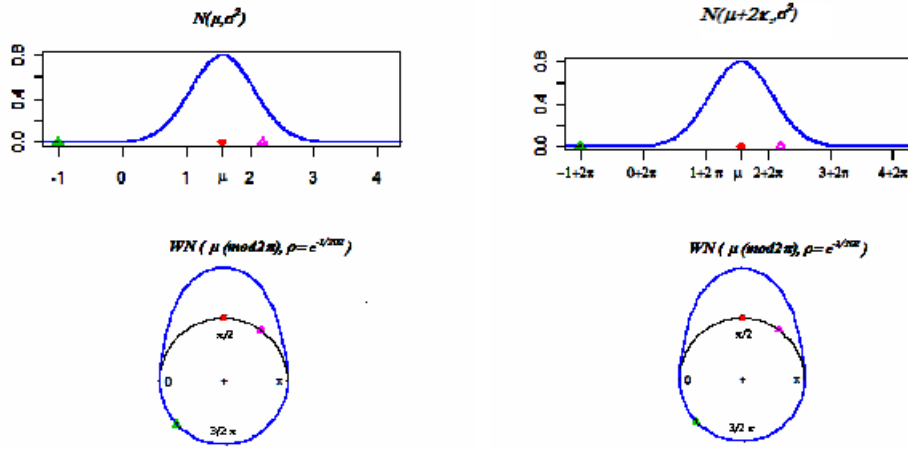


Figure 3.12: Normal density and its corresponding Wrapped Normal density with $\mu = \frac{\pi}{2}$ (left) and $\mu = \frac{\pi}{2} + 2\pi$ (right)

3.4 Identifiability Problem

After detecting the kind of data sets where the inference procedure fails, we now investigate the role of the wrapping coefficients k 's in the inference procedure, first in general and then in the particular case of data sets with large variability.

3.4.1 Wrapped density approximation

Recalling the definition given in Section 2.4.3 about the circular variable X and its probability density function

$$f_w(x) = \sum_{k=-\infty}^{\infty} f(x + 2k\pi), \quad 0 \leq x < 2\pi,$$

we note that the k 's coefficients have to be interpreted as the “tool” to expand the circular domain $[0, 2\pi)$ to the in line domain \mathbb{R} , so that, an in line density distribution f can be applied to the argument $x + 2k\pi$. Then, the unobserved values k 's determine the support on which the in line density is fitted and, in this sense, the central role of these coefficients in the inference parameter estimation is easily realizing.

To understand the meaning of the k 's, the graphical representation of the wrapped Normal distributions shown in Figure 3.12 can be useful. Examining this plot we realize that the same wrapping distribution can be obtained from different in line distributions. More precisely, the same wrapping Normal distribution can be obtained from different in line distributions that have the same scale but a different location parameter, i.e having $2k\pi$ shifted mean. Therefore, by only observing the circular distribution it is difficult to identify which is the in line distribution that have generated it, i.e. the k 's values are not uniquely identified.

Basically, the point that we want to investigate is summarized by the following

$Y \sim N(0, \pi/2)$ $X=Y \bmod(2\pi)$	x						
	$-\pi$	$-2\pi/3$	$-\pi/3$	0	$\pi/3$	$2\pi/3$	π
$Pr(Y = x \mid X = x) \text{ (k=0)}$	0.500	0.985	1.000	1.000	1.000	0.985	0.500
$Pr(Y = x + 2\pi \mid X = x) \text{ (k=1)}$	0.500	0.015	0.000	0.000	0.000	0.000	0.000
$Pr(Y = x - 2\pi \mid X = x) \text{ (k=-1)}$	0.000	0.000	0.000	0.000	0.000	0.015	0.500

Table 3.10: Conditional probabilities using $Y \sim N(0, \pi/2)$

sentence: “given a circular observed value x , drawn for example from a wrapped Normal distribution, which is the most likely corresponding in line value y ?” To answer to this question, we extend a Mardia’s assertion (Mardia and Jupp, 2000, p. 50) saying that when the in line Normal distribution variance is less than 2π , the term with $k = 0$ of the density (2.43) gives a reasonable approximation of the same density. In particular, we approximate the wrapped density distribution (2.43) with the terms $k = 0$, $k = 1$ and $k = -1$ and, for each of them, we compute the probability $Pr(Y \mid X = x)$ in order to determine the most likely in line values for a given x .

Suppose to have the in line Normal distribution $Y \sim N(0, \sigma^2)$, the density approximation with the terms $k = 0$, $k = 1$, $k = -1$ of the wrapped Normal distribution is given by:

$$f(x) = (2\pi\sigma^2)^{-1/2} \left[\exp \left\{ -\frac{x^2}{2\sigma^2} \right\} + \exp \left\{ -\frac{(x + 2\pi)^2}{2\sigma^2} \right\} + \exp \left\{ -\frac{(x - 2\pi)^2}{2\sigma^2} \right\} \right], \quad (3.13)$$

where the first term corresponds to no wrapping, the second to a positive wrapping and the third to a negative wrapping.

The target probability $Pr(Y \mid X = x)$ will be $Pr(Y = x \mid X = x)$ if we use the approximation with $k = 0$; $Pr(Y = x + 2k\pi \mid X = x)$ if we use the approximation with $k = 1$ and $Pr(Y = x - 2k\pi \mid X = x)$ if we use $k = -1$. Moreover, $Pr(Y = x \mid X = x) = \frac{Pr(Y=x, X=x)}{Pr(X=x)}$, where $Pr(Y = x, X = x)$ is the first term of (3.13) while $Pr(X = x)$ coincides with (3.13). In the same way, $Pr(Y = x + 2\pi, X = x)$ is the second term of the (3.13) and $Pr(Y = x - 2\pi, X = x)$ is the third term.

As an example we apply the density approximation to four wrapped Normal densities obtained wrapping the in line distributions $Y \sim N(0, \pi/2)$, $Y \sim N(0, \pi)$, $Y \sim N(0, 4\pi/3)$, $Y \sim N(0, 2\pi)$, respectively. The results, reported in Tables 3.10- 3.13 and in Figure 3.13, represent the probabilities that an observed circular value x is obtained by wrapping the in line value y k -times around the circle, with $k \in \{-1, 0, 1\}$.

Observing Table 3.10 and Figure 3.13(a) we can see that the value $2\pi/3 = 120^\circ$ has a probability of 0.015 to come from the in line distribution represented by the approximation with $k = -1$ while the equivalent probability with $k = 0$ is 0.985. Table 3.10 also shows that any wrapping of an order greater than 1 is unlikely to occur if the in line variance takes a value of $\pi/2$. Moreover, it easy to calculate the

$Y \sim N(0, \pi)$ $X=Y \bmod(2\pi)$	x						
	$-\pi$	$-2\pi/3$	$-\pi/3$	0	$\pi/3$	$2\pi/3$	π
$Pr(Y = x \mid X = x) \text{ (k=0)}$	0.500	0.890	0.985	0.996	0.985	0.890	0.500
$Pr(Y = x + 2\pi \mid X = x) \text{ (k=1)}$	0.500	0.110	0.015	0.002	0.000	0.000	0.000
$Pr(Y = x - 2\pi \mid X = x) \text{ (k=-1)}$	0.000	0.000	0.000	0.002	0.015	0.110	0.500

Table 3.11: Conditional probabilities using $Y \sim N(0, \pi)$

$Y \sim N(0, 4\pi/3)$ $X=Y \bmod(2\pi)$	x						
	$-\pi$	$-2\pi/3$	$-\pi/3$	0	$\pi/3$	$2\pi/3$	π
$Pr(Y = x \mid X = x) \text{ (k=0)}$	0.500	0.828	0.957	0.982	0.957	0.828	0.500
$Pr(Y = x + 2\pi \mid X = x) \text{ (k=1)}$	0.500	0.172	0.041	0.009	0.002	0.000	0.000
$Pr(Y = x - 2\pi \mid X = x) \text{ (k=-1)}$	0.000	0.000	0.002	0.009	0.041	0.172	0.500

Table 3.12: Conditional probabilities using $Y \sim N(0, 4\pi/3)$

$Y \sim N(0, 2\pi)$ $X=Y \bmod(2\pi)$	x						
	$-\pi$	$-2\pi/3$	$-\pi/3$	0	$\pi/3$	$2\pi/3$	π
$Pr(Y = x \mid X = x) \text{ (k=0)}$	0.500	0.737	0.878	0.920	0.878	0.737	0.500
$Pr(Y = x + 2\pi \mid X = x) \text{ (k=1)}$	0.500	0.259	0.108	0.040	0.013	0.004	0.001
$Pr(Y = x - 2\pi \mid X = x) \text{ (k=-1)}$	0.001	0.004	0.013	0.040	0.108	0.259	0.500

Table 3.13: Conditional probabilities using $Y \sim N(0, 2\pi)$

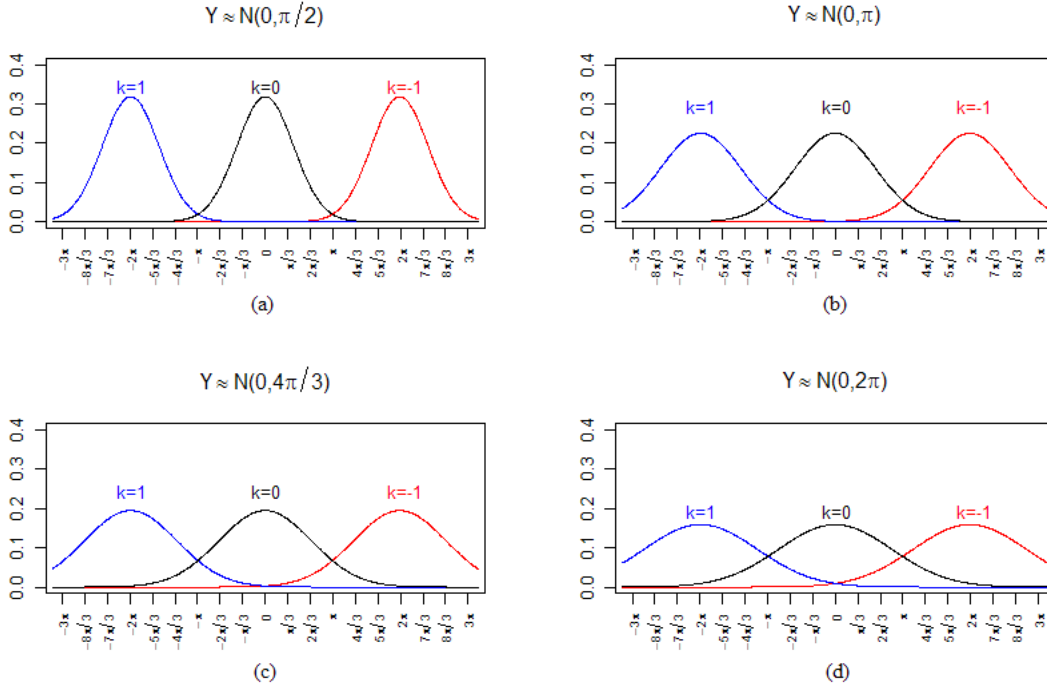


Figure 3.13: Probability density plots of Normal distributions conditioned on wrapped values $X = x$, related to Tables 3.10-3.13

probability that the wrapping of Y has no effect, that is $Y = x$, as the probability of the Normal distribution calculate on the support $[-\pi, \pi)$. The probability values for the four distributions described in Tables 3.10-3.13 are listed in Table 3.14.

It is interesting to note that all these probabilities are high although they decrease when the in line, and the corresponding wrapped, variance increase. This last issue is evident if we look at Figure 3.13. In particular, comparing plot (a) and (d) for $x = 2\pi/3$, we realize that when the variance is small (case (a)) x has a very small probability of coming from an in line distributions represented by $k \neq 0$; when the variance increases till 2π (case (d)), instead, x has a not negligible probability of coming from the in line distributions represented by $k \neq 0$.

$Y \sim N(0, \sigma^2)$	
σ^2	$Pr(Y = x)$
$\pi/2$	0.988
π	0.924
$4\pi/3$	0.875
2π	0.790

Table 3.14: Probability that the wrapping has no effect ($Pr(Y = x)$) related to four Normal distributions with zero means and different variances listed in Tables 3.10-3.13

What we have described constitute the *identifiability problem on k 's* that occurs when we have to handle data with large variability. Moreover, we have verified the close relationship between the distribution variance and k identifiability problem and, then, demonstrate the reason why the inference procedure fails when the data exhibit large variance. Finally, this study on identifiability problem for k reveals that the density approximation with the three terms $k = -1$, $k = 0$ and $k = 1$ represents a good approximation for distributions with variance until 2π and higher; allowing us to use this information in the parameter estimation procedure, introducing for example a prior distribution on wrapping coefficients k 's directly. Recalling the posterior density (3.10) in Section 3.1 and collecting the k 's in the vector \mathbf{K} , we can rewrite it as:

$$\begin{aligned} Pr(\mathbf{K} \mid \mathbf{X}, \Psi) &= \frac{Pr(\mathbf{K}, \mathbf{X}, \Psi)}{Pr(\mathbf{X}, \Psi)} = \frac{Pr(\mathbf{X} \mid \Psi, \mathbf{K})Pr(\mathbf{K})}{Pr(\mathbf{X}, \Psi)} \\ &\propto Pr(\mathbf{X} \mid \Psi, \mathbf{K})Pr(\Psi \mid \mathbf{K})Pr(\mathbf{K}), \end{aligned} \quad (3.14)$$

where $Pr(\mathbf{X} \mid \Psi, \mathbf{K})$ is still distributed as multivariate Normal density while through the $Pr(\mathbf{K})$ we can give a prior probability to the wrapping coefficients. A reasonable one is the Multinomial distribution on values $\{-1, 0, 1\}$. In particular, by the simulation-study of this section (see Figure 3.13 and related tables) we realize that when the circular mean is between 0° and π , the two most likely values for k 's are 0 and -1 ; whereas when the circular mean is between π and 2π the most likely values are 0 and 1.

At this point, a legitimate objection can be raised: the wrapping approach applicability is limited to small variability data sets because of the k 's identifiability problem described above. Now we show that this is a misrepresented problem in that for variance over a “threshold”, the wrapped Normal distribution can be approximated by the circular Uniform distribution. In this case there is no reason to use the wrapped Normal distribution to fit the data. Next, we aim to find an analytic derivation of a “threshold” value for the variance.

3.4.2 Tests of uniformity and threshold value for variance

Here we present the main tests for uniformity in order to assess when a wrapped Normal distribution can be approximated by a circular Uniform distribution. We first give a formal description of these tests and then we apply them to different simulated data sets.

Rayleigh's test

The Rayleigh's test is based on the intuitive idea of rejecting uniformity when the vector sample mean (\bar{C}, \bar{S}) , defined in Section 2.2.1, is far from 0, i.e., recalling equations (2.2) and (2.3), when \bar{R} is large.

The Rayleigh's test is the score test of uniformity within the von Mises model

(2.36). Put $\omega = (h \cos \mu, h \sin \mu)^T$, the log-likelihood based on circular observations x_1, \dots, x_n is

$$l(\omega; x_1, \dots, x_n) = n\omega^T \bar{\mathbf{x}} - n \log I_0(h),$$

where $I_0(h)$ is the modified Bessel function and $\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n (\cos x_i, \sin x_i)^T$ is the sample mean vector.

The score is:

$$\mathbf{U} = \frac{\partial l}{\partial \omega^T} = n\bar{\mathbf{x}} - nA(h)(\cos \mu, \sin \mu)^T. \quad (3.15)$$

From the moments properties of \bar{R} , it is possible to demonstrate (Mardia and Jupp, (2000), pp. 94-95) that the score statistic is

$$\mathbf{U}^T \text{var}(\mathbf{U})^{-1} \mathbf{U} = 2n\bar{R}^2 \quad (3.16)$$

From the general theory of score test (see Cox and Hinkley, 1974), the large sample asymptotic distribution of $2n\bar{R}^2$ under uniformity is a Chi-square with two degrees of freedom:

$$2n\bar{R}^2 \sim \chi_2^2, \quad (3.17)$$

where n is the sample size. It has been demonstrated, also, that the Rayleigh's test coincides with the likelihood test of uniformity within the von Mises family.

Kuiper's test

The Kuiper's test is based on the empirical distribution function. As for in line case, this test measures the deviation between the empirical distribution, $S_n(x)$, and the Uniform cumulative distribution functions (cdf), $F(x) = x/2\pi$. In the case of circular data, the definition of cumulative distribution function is not obvious and is quite different from the in line cdf. In the circular data case, in fact, we first have to choose the circle zero point and orientation, then we need to augment the ordered observations, x_1, \dots, x_n , of $x_{(0)} = 0$ and $x_{(n+1)} = 2\pi$. The S_n is then defined by:

$$S_n(x) = \frac{i}{n} \quad \text{if } x_{(i)} \leq x \leq x_{(i+1)} \quad i = 0, 1, \dots, n. \quad (3.18)$$

Just as in Kolmogorov-Smirnov's test for in line distribution (see Durbin, 1973), the following quantities are defined:

$$D_n^+ = \sup_x \{S_n(x) - F(x)\}, \quad D_n^- = \sup_x \{F(x) - S_n(x)\}.$$

To overcome the dependence of D_n^+ and D_n^- on the choice of the initial direction, Kuiper (1960) defined

$$V_n = D_n^+ + D_n^-. \quad (3.19)$$

The statistic (3.19) has been demonstrated (see Mardia and Jupp, 2000, p.101; Jammalamadaka and SenGupta, 2001, pp.154-155) to be invariant under the change of initial direction. The null hypothesis of uniformity is rejected for large values of

V_n . Moreover, the Kuiper's test is consistent against all alternative to uniformity. For practical purposes, the following modification of V_n is used:

$$V_n^* = n^{1/2} V_n \left(1 + \frac{0.155}{\sqrt{n}} + \frac{0.24}{n} \right).$$

Table of upper quantiles of V_n^* are given in Stephens (1970) and Arsham (1988).

Watson's test

Watson (1961) provided a modification of the Cramér-von Mises test (see Durbin, 1973) suitable for circular data. The Watson's statistics is:

$$W_n^2 = \int_0^{2\pi} \left[(S_n(x) - F(x)) - \int_0^{2\pi} (S_n(x) - F(x)) dF \right]^2 dF. \quad (3.20)$$

It follows from this definition that the Watson's statistic is invariant under rotations and reflections.

As for the Kuiper's test, it is useful to consider the following modified statistic:

$$W_n^{*2} = \left(W_n^2 - \frac{0.1}{n} + \frac{0.1}{n^2} \right) \left(1 + \frac{0.8}{n} \right).$$

The upper quantiles of W_n^{*2} were calculated by Stephens (1970).

Rao's spacing test

The next is one of the so called *spacing tests* because they are based on the sample arc lengths T_1, \dots, T_n defined as:

$$T_i = x_{(i)} - x_{(i-1)}, \quad i = 1, \dots, n-1, \quad T_n = 2\pi - (x_{(n)} - x_{(1)}).$$

Under uniformity $E[T_i] = 2\pi/n$. Hence, it is reasonable to reject uniformity for large values of

$$L = \frac{1}{2} \sum_{i=1}^n \left| T_i - \frac{2\pi}{n} \right|. \quad (3.21)$$

Large values of L indicate clustering of observations. This statistic was introduced by Rao (1969) and extends to circular data the corresponding in line test suggested by Kendall (1946).

An extensive table of quantiles of L is given in Russell and Levitin (1996), while Sherman (1950) shows that a suitable transformation of L is asymptotically standard Normal distributed.

Here we apply the four uniformity tests to several simulated wrapped Normal data sets that differ in sample size and variability. Results are listed in Tables 3.15-3.17.

When the sample size is equal to 30, the tests reveal uniformity even with a variance larger than $\pi/2$ (in the simulated example with variance equal to $2\pi/3$), while when the sample size increases to 100 and 1000, the uniformity is detected with variances

n=30	Rayleigh		Kuiper		Watson		Rao Spacing	
X=Y(mod 2pi)	Stat.	p-value	Stat.	p-value	Stat.	p-value	Stat.	p-value
$X \sim WN(0, \pi/2)$	0.43	0.003	2.08	< 0.01	0.31	< 0.01	157.94	< 0.05
$X \sim \mathbf{WN}(\mathbf{0}, \mathbf{2\pi/3})$	0.29	0.08	1.56	> 0.15	0.14	> 0.10	126.04	> 0.01
$X \sim WN(0, 5\pi/6)$	0.19	0.34	1.24	> 0.15	0.08	> 0.10	125.39	> 0.10
$X \sim WN(0, \pi)$	0.23	0.21	1.30	> 0.15	0.10	> 0.10	124.74	> 0.10
$X \sim WN(0, 7\pi/6)$	0.17	0.43	1.29	> 0.15	0.08	> 0.10	127.30	> 0.10
$X \sim WN(0, 4\pi/3)$	0.18	0.38	1.24	> 0.15	0.06	> 0.10	124.18	> 0.10
$X \sim WN(0, 5\pi/3)$	0.21	0.24	1.42	> 0.15	0.09	> 0.10	138.26	> 0.10
$X \sim WN(0, 2\pi)$	0.16	0.47	1.13	> 0.15	0.07	> 0.10	130.40	> 0.10
$X \sim WN(0, 3\pi)$	0.12	0.67	0.94	> 0.15	0.04	> 0.10	130.23	> 0.10
$X \sim WN(0, 10\pi)$	0.09	0.75	0.90	> 0.15	0.04	> 0.10	125.65	> 0.10

Table 3.15: Tests of uniformity for wrapped Normal density with sample size $n = 30$. The bold line indicates the first variance value beyond which the test accepts the hypothesis of uniformity.

n=100	Rayleigh		Kuiper		Watson		Rao Spacing	
X=Y(mod 2pi)	Stat.	p-value	Stat.	p-value	Stat.	p-value	Stat.	p-value
$X \sim WN(0, \pi/2)$	0.52	$2e - 8$	3.64	< 0.01	1.37	< 0.01	162.53	< 0.001
$X \sim WN(0, 2\pi/3)$	0.38	$7e - 7$	8.04	< 0.01	7.18	< 0.01	153.24	< 0.001
$X \sim WN(0, 5\pi/6)$	0.32	$2e - 5$	2.53	< 0.01	0.54	< 0.01	142.39	> 0.10
$X \sim WN(0, \pi)$	0.28	$3e - 4$	2.37	< 0.01	0.45	< 0.01	132.78	> 0.10
$X \sim \mathbf{WN}(\mathbf{0}, \mathbf{7\pi/6})$	0.11	0.27	1.35	> 0.15	0.09	> 0.10	125.95	> 0.10
$X \sim WN(0, 4\pi/3)$	0.10	0.35	1.22	> 0.15	0.09	> 0.10	123.48	> 0.10
$X \sim WN(0, 5\pi/3)$	0.11	0.24	1.42	> 0.15	0.1	> 0.10	125.26	> 0.10
$X \sim WN(0, 2\pi)$	0.10	0.32	1.03	> 0.15	0.07	> 0.10	126.10	> 0.10
$X \sim WN(0, 3\pi)$	0.05	0.74	1.19	> 0.15	0.07	> 0.10	122.65	> 0.10
$X \sim WN(0, 10\pi)$	0.08	0.50	1.15	> 0.15	0.06	> 0.10	122.60	> 0.10

Table 3.16: Tests of Uniformity for wrapped Normal density with sample size $n = 100$. The bold line indicates the first variance value beyond which the test accepts the hypothesis of uniformity.

n=1000	Rayleigh		Kuiper		Watson		Rao Spacing	
$X=Y(\bmod 2\pi)$	Stat.	p-value	Stat.	p-value	Stat.	p-value	Stat.	p-value
$X \sim WN(0, \pi/2)$	0.47	$8e - 18$	10.10	< 0.01	11.53	< 0.01	165.47	< 0.001
$X \sim WN(0, 2\pi/3)$	0.36	$6e - 15$	23.13	< 0.01	65.79	< 0.01	150.45	< 0.001
$X \sim WN(0, 5\pi/6)$	0.29	$5e - 12$	6.14	< 0.01	4.42	< 0.01	145.93	< 0.001
$X \sim WN(0, \pi)$	0.21	$3e - 9$	4.58	< 0.01	2.29	< 0.01	134.36	> 0.10
$X \sim WN(0, 7\pi/6)$	0.17	$8e - 7$	3.73	< 0.01	1.45	< 0.01	135.02	> 0.10
$X \sim WN(0, 4\pi/3)$	0.12	$2e - 5$	3.02	< 0.01	0.82	< 0.01	135.60	> 0.10
$X \sim WN(0, 5\pi/3)$	0.09	$3e - 4$	2.43	< 0.01	0.44	< 0.01	134.50	> 0.10
$X \sim WN(0, 2\pi)$	0.06	0.027	1.97	< 0.025	0.24	< 0.025	132.74	> 0.10
$X \sim WN(0, 3\pi)$	0.03	0.46	1.15	> 0.15	0.07	> 0.10	134.10	> 0.10
$X \sim WN(0, 10\pi)$	0.02	0.64	1.22	> 0.15	0.06	> 0.10	129.90	> 0.10

Table 3.17: Tests of Uniformity for wrapped Normal density with sample size $n = 1000$. The bold line indicates the first variance value beyond which the test accepts the hypothesis of uniformity.

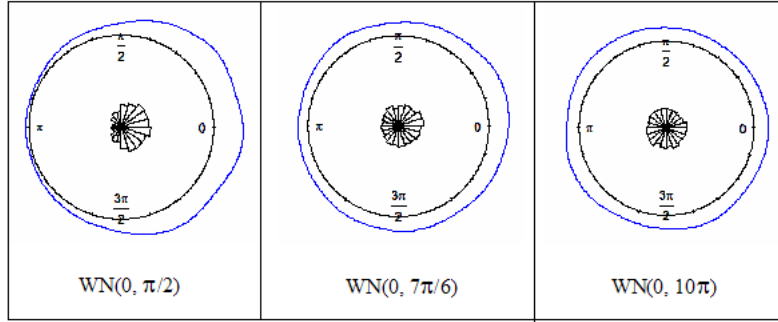


Figure 3.14: Plots of wrapped Normal distributions ($n=100$) with different variances.

larger than π (i.e. with variance $7\pi/6$) and 2π (i.e. with variance 3π) respectively. To realize what happens when the in line and the corresponding circular variance increase, the plot depicted in Figure 3.14, related to the case of $n = 100$, can be useful. We note that when the variance is larger than π (central and right plots) there is no different in practice between the wrapped Normal distribution and the circular Uniform distribution.

Nevertheless, the above results are only empirical. Here we want to derive analytically a variance threshold over which we can approximate the wrapped Normal distribution to the Uniform distribution. For this purpose we remind the quantity that the Rayleigh test is based on, i.e. the sample mean resultant length \bar{R} . Its corresponding population quantity is given by ρ , that is the concentration wrapped Normal parameter and a transformation of the in line Gaussian distribution variance, $\rho = e^{-\sigma^2/2}$. Moreover, we know from the (ii) property (see Section 2.4.2) of

the von Mises distribution that

$$\rho = A(h)$$

with

$$A(h) = J_1(h)/J_0(h), \quad (3.22)$$

where $J_p(h)$ is the modified Bessel function of order p defined in (2.40).

So, both the wrapped Normal variance (or concentration) and the Rayleigh's test statistic are based on $A(h)$ quantity. Here the point is: the function $J_p(h)$ has two power series expansions, one for *small* h and one for *large* h . Consequently the function $A(h)$ has two series expansions for *small* h and for *large* h , too. In particular, for small h it is:

$$A_s(h) = \frac{h}{2} \left\{ 1 - \frac{1}{8}h^2 + \frac{1}{48}h^4 + o(h^4) \right\}, \quad (3.23)$$

while for large h it is:

$$A_l(h) = 1 - \frac{1}{2h} - \frac{1}{8h^2} - \frac{1}{8h^3} + o(h^{-3}). \quad (3.24)$$

The idea for determining the variance threshold is to find the value of h for which the two power series expansions (3.23) and (3.24) are 'close' each other and to $A(h)$. Equalizing the ratio of the first terms of both series expansions (3.23) and (3.24) to one we obtain the following expression:

$$\begin{aligned} \frac{A_s(h)}{A_l(h)} &= \frac{\frac{h}{2} + o(h)}{1 - \frac{1}{2h} + o(h^{-1})} \\ &= \frac{h^2 + o(h^2)}{2h - 1 + o(1)} = 1 \\ &\Leftrightarrow h^2 - 2h + 1 + o(1) = 0 \\ &\Leftrightarrow h^2 - 2h + 1 + O(h) = 0 \\ &\Rightarrow h = 1 + O(h) \end{aligned} \quad (3.25)$$

The result (3.25) indicates that the threshold we are searching is in a neighborhood of 1 of length 1, i.e. it is in the set $[0, 2]$. In order to have a more precise value, we compute, for different values of h , the numerical approximations of both quantities (3.23) and (3.24) and, using the relationship $\rho = e^{-\sigma^2/2}$, the corresponding variances σ_s^2 and σ_l^2 . The results are reported in Table 3.18.

Comparing the values of $A_s(h)$ and $A_l(h)$ to the tabulated values of the function $A(h)$ based on Table C of Batschelet (1965), we realize that the approximation is based on $A_s(h)$ for the values of h in $(0, 1.4]$ corresponding to variance σ^2 in $[1.074, \infty)$. Whereas, for larger values of h , corresponding to $\sigma^2 \leq 1.109$, the approximation based on $A_l(h)$ appears more suitable than $A_s(h)$. This leads to establish threshold values for the variance that can be summarized as the follows:

h	$A_s(h)$	$A_l(h)$	$A(h)$	σ_s^2	σ_l^2
0.0	0.000	0.000	0.000	∞	-
0.1	0.050	-141.5	0.050	5.991	-
0.2	0.100	-20.25	0.100	4.605	-
0.3	0.148	-6.685	0.148	3.817	-
0.4	0.196	-2.984	0.196	3.258	-
0.5	0.242	-1.500	0.242	2.833	-
0.6	0.287	-0.759	0.287	2.494	-
0.7	0.330	-0.334	0.330	2.215	-
0.8	0.371	-0.064	0.371	1.981	-
0.9	0.410	0.119	0.410	1.780	4.263
1.0	0.446	0.250	0.446	1.606	2.773
1.1	0.481	0.348	0.481	1.453	2.109
1.2	0.513	0.424	0.513	1.316	1.715
1.3	0.543	0.485	0.543	1.191	1.449
1.4	0.575	0.534	0.570	1.074	1.256
1.5	0.618	0.580	0.596	0.962	1.109
1.6	0.653	0.620	0.620	0.852	0.994
1.7	0.691	0.640	0.640	0.740	0.901
1.8	0.732	0.662	0.662	0.623	0.824
1.9	0.779	0.681	0.681	0.499	0.759
2.0	0.833	0.698	0.698	0.365	0.704
2.1	0.897	0.714	0.714	0.218	0.657
2.2	0.971	0.728	0.728	0.058	0.615
2.3	1.060	0.741	0.741	-	0.579
2.4	1.165	0.754	0.754	-	0.546446
2.5	1.291	0.765	0.765	-	0.517541

Table 3.18: Numerically approximations of $A(h)$ function and corresponding variances.

- $h < 1 \Leftrightarrow \sigma^2 > \pi/2$: the variance is considered "large";
- $1 \leq h \leq 2 \Leftrightarrow 0.72 \leq \sigma^2 \leq \pi/2$: the variance has intermediate values and both the approximation $A_s(h)$ and $A_l(h)$ can be used;
- $h > 2 \Leftrightarrow \sigma^2 < 0.72$: the variance is considered "small".

These variance thresholds are coherent with the empirical results obtained by the simulation-base study on test of uniformity. So, we can assert that when the in line variance, and its corresponding circular, are larger than $\pi/2$ the wrapped Normal circular distribution can be well approximated by Uniform circular distribution.

3.5 Discussion

In this chapter we illustrate the MCMC estimate procedure for the wrapping approach. Through univariate and multivariate simulated examples, we point out some inference problems that occur in specific kinds of data; in particular the data variability effect on inference estimate results is studied. We find a strong relationship between 'large' data variability and poor inference results due to the identifiability problem on wrapping coefficients (see Section 3.3).

We investigate and deeply analyze the wrapping coefficients identifiability problem, providing original evidences and results. From this study, we realize two key aspects of both circular data and the wrapping approach. We demonstrate that the suitable distribution for circular data with 'large' variability is the circular Uniform distribution. In particular, we provide an analytic value-definition of 'large' variance, equal to $\pi/2$, beside that the circular distribution can be well approximated by circular Uniform distribution. Moreover, we numerically show that this threshold value increases until π when the sample size is $n = 100$.

The second key aspect is related to wrapped circular distribution. We demonstrate that when the data variability is small, the wrapped Normal density $f(x) = (2\pi\sigma^2)^{-1/2} \sum_{k=-\infty}^{\infty} \exp\left\{-\frac{(x-\mu+2k\pi)^2}{2\sigma^2}\right\}$ is adequately approximated by only three k -values $\{-1, 0, 1\}$. This evidence, here originally showed, is directly used into the Bayesian inference procedure, through an opportune prior probability on k coefficients, solving the identifiability problem and allowing a more efficient inference on the wrapped distribution model.

Thus, the opportunity of overcoming the difficulty due to the identifiability wrapping coefficients problem, allows us to appreciate the main advantages of the wrapping approach: the *ease of extending to the multivariate case* and the *easy interpretability of the parameters* in terms of circular phenomena behavior. With regard to these advantages, we point out that the examples in literature in which the wrapping approach is used concern only with a bivariate wrapped Normal distribution (see e.g. Coles, 1998) and a univariate regression model and univariate autoregressive model (see for example Ravindran, 2002), i.e. regard only computationally simple models. In the same way, we point out also that in the most recent works concerning the von Mises distribution, the attempts of extending to multivariate case, Mardia

et al. (2007) and Mardia *et al.* (2008), present difficulties in both definition and estimation of probability models even with low model dimension. Thus, in case of multivariate framework and, more in general, with complex models, the wrapping Normal is preferable to von Mises distribution.

In the next chapters we show that the new inference procedure setting, that involves the prior on k , allows to implement fairly complicated models such as the measurement error model in a multivariate framework.

Chapter 4

Measurement Error Model

In this chapter we deal with measurement error (ME) models, also called error-in-variables models. Here we give a general overview of these models, while the main applications and discussions about it will be in the next chapters. In particular we refer to ME models in the spatiotemporal models context.

For the simplest linear ME model from bivariate data, the goal is to estimate a straight line fit between the two variables, both of which are measured with error. Adcock (1877) is usually recorded as the first scholar who specifically considers such model.

In this chapter we first introduce the ME model and its properties in a *in line* standard context. Then, we derive the *wrapped* version of the ME model and analyze, also in this case, the related properties.

4.1 ME Models Definition

In this section we give the definition and a brief overview of the main properties and inference methods regarding the ME models. A comprehensive account of linear measurement error models can be found in Fuller (1987). More recently, discussion on statistical regression with measurement error is available in Cheng and Van Ness (1999); whereas useful extension to ME in nonlinear models are discussed in Carroll *et al.* (1995), where it is possible to find a brief reference to the Bayesian approach for the ME models, also. A very recent work on ME model applied to dental data is in Ghosh *et al.* (2008). Moreover, a paper on a fully Bayesian approach to nonparametric regression problem with error in variables is presented in Berry *et al.* (2002).

Classical ME Model

The standard regression model with one explanatory variable is

$$\nu = \beta_0 + \beta_1\chi + \xi \tag{4.1}$$

where the independent variable, χ , is either fixed or random and the error, ξ , has zero mean and is uncorrelated with χ .

The corresponding *classical linear ME* model assumes that the variable χ and γ are related by

$$\gamma = \beta_0 + \beta_1\chi \quad (4.2)$$

but that the “true” variables χ and γ are unobservable and can only be observed with additive errors. Thus, instead of observing χ and γ directly, we observe the variables

$$\omega = \chi + \epsilon \quad \text{and} \quad \nu = \gamma + \xi \quad (4.3)$$

where χ and the errors, ϵ and ξ , are uncorrelated.

Because both χ and γ are measured with error, model (4.2)-(4.3) is called *measurement error model*.

Applications in which the variable χ is measured with error are perhaps more common than those in which χ is measured precisely. Most medical variables as well as variables observed in biological or physical sciences are measured with non-negligible error.

For a sample of size n , the linear univariate ME model (4.2)-(4.3) can be formulated as follows. The unobserved variables (χ_i, γ_i) satisfy

$$\gamma_i = \beta_0 + \beta_1\chi_i, \quad i = 1, 2, \dots, n. \quad (4.4)$$

However, we observe (ω_i, ν_i) which are the observed variables plus the additive errors (ϵ_i, ξ_i) :

$$\omega_i = \chi_i + \epsilon_i \quad \text{and} \quad \nu_i = \gamma_i + \xi_i \quad i = 1, 2, \dots, n. \quad (4.5)$$

There are two separate models for the form (4.2)-(4.3) depending on the assumption about χ_i . If the χ 's are fixed (or random but in latter case no or at least only minimal assumptions are made about the distribution of the χ_i), then the model is known as a *functional model*; whereas if χ 's are independent identically distributed random variables and independent of (not just uncorrelated with) the errors, the model is known as *structural model*, (Fuller, 1987).

Berkson Model

An important special case of ME models is the *Berkson* model. Berkson (1950) proposed a different regression model with measurement error which is appropriate in many applications. A key assumption in the classical ME model is that in equations (4.5) the true variable χ_i is independent of the errors (ϵ_i, ξ_i) . Berkson called the ω_i under this assumption, uncontrolled observation. If the measured quantity is fixed but the true value differs by an error from the measurement, then

$$\chi_i = \omega_i + \epsilon_i, \quad (4.6)$$

that is, the true value is the observed value plus a random error. Berkson called such ω_i controlled observation.

The main consequence in writing the error model as (4.6) rather than as (4.5) is that in the classical ME model the conditional distribution of ω given χ is modeled; while in the Berkson model the conditional distribution of χ given ω is modeled. This corresponds to the fact that χ is independent of ϵ in the classical ME model whereas ω is independent of ϵ in the Berkson model.

4.2 Properties and Parameter Estimation Methods for ME Models

Although the regression model can be considered a special case of ME models, here it is useful to point out some properties that distinguish ME models from its special case.

One of the most important difference between ME models and ordinary regression models concerns model identifiability. It is often assumed that all the random variables in the ME model are jointly Normal distributed. In this case, model (4.2)-(4.3) is a not identifiable structural model. This means that different sets of parameters can lead to the same joint distribution of ω and ν , and in this situation it is impossible to estimate consistently the parameters from data without further assumptions. There are six *side conditions* any one of which makes the Normal structural model identifiable:

- (a) the ratio of the error variance, $\lambda = \sigma_\xi^2/\sigma_\epsilon^2$, is known;
- (b) the χ reliability ratio, k_χ , is known;
- (c) σ_ϵ^2 is known;
- (d) σ_ξ^2 is known;
- (e) both of the error variance, σ_ξ^2 and σ_ϵ^2 , are known;
- (f) the intercept, β_0 , is known and $\mathbb{E}(\chi) \neq 0$.

Assumption (a) is the most popular of these additional assumptions and is the one with the most published theoretical results. It has a long history dating back to Adcock (1877). Assumption (b) is commonly found in social sciences and psychology literature. The ratio

$$k_\chi = \frac{\sigma_\chi^2}{\sigma_\chi^2 + \sigma_\epsilon^2} = \frac{Var(\chi)}{Var(\omega)},$$

is called the *reliability ratio* (Fuller, 1987, Section 1.1) and is related to the bias of the estimated slope parameter β_1 of the standard regression model. In the ME model, the predictor χ cannot be observed directly, but we observe $\omega = \chi + \epsilon$ where ϵ is independent of χ , has zero mean and variance σ_ϵ^2 . In this case the ordinary least squares regression of ν on ω is a consistent estimate not of β_1 but instead of $\beta_{1*} = k_\chi \beta_1$. Thus, the least squares regression of ν on ω produces an estimator that is biased. The bias is attenuated toward zero but increase as σ_ϵ^2 increases. This is one of the more important differences between the ME model and the standard regression model. A general overview about bias and attenuation issues can be founded in Fuller (1987) and in Carroll *et al.* (1995).

Assumption (c) has gained attention recently and is a popular assumption for non-linear models (see Cheng and Van Ness, 1999, Chap. 6). Moreover, it is also likely to be realistic in many applications because it is not uncommon to have replications

of ω , which can be used to estimate σ_ϵ^2 . Assumption (d) and (f), instead, are less useful and cannot be used to make the structural ME model with more than one explanatory ω variable identifiable. Finally, assumption (e) frequently leads to the same estimates as those for assumption (a).

For the *functional* ME model, even without any side conditions, the parameters β_0 and β_1 are identifiable (Bowden and Turkington, 1984); but, unfortunately, this does not guarantee the existence of a consistent estimate of β_1 , thus, also for the functional model we need some side conditions or other additional information to obtain consistent estimates. For the functional model, then, side conditions (a) and (e) can be used to obtain consistent maximum likelihood estimates. Whereas, the consistent estimates resulting from the other structural model side conditions do not come from maximum likelihood. Finally, the condition (b) is not well defined for the functional model. A comprehensive account of maximum likelihood parameter estimate, by using the side conditions, for both structural and functional models can be found in Fuller (1987) and Cheng and Van Ness (1999).

The side conditions are needed to estimate the Berkson model parameters as well, (see Fuller, 1987). Whereas it is demonstrated (Cheng, 1994), that the least squares estimates for Berkson model coincide with the usual regression least squares estimates.

An alternative approach for estimating the parameters of a ME model uses *instrumental variables (IV)*. Instrumental variables are adopted in place of side conditions to make the model identifiable. As a matter of fact, in absence of information about the measurement error variance, the ME model parameter estimation is possible only if other variables (instrumental variables) are provided in addition to the unbiased measurement $\omega = \chi + \epsilon$. One potential source of an instrumental variables is a second measurement of χ obtained by an independent method. Consider the ordinary regression model with one explanatory variable ω ,

$$\nu_i = \beta_0 + \beta_1 \omega_i + \epsilon_i \quad (4.7)$$

with $Cov((\epsilon_1, \dots, \epsilon_n)') = \sigma_\epsilon^2 \mathbf{I}$ and the ω_i independent and identically distributed. Suppose, in this situation, that a random sample $z = (z_1, \dots, z_n)$ is available such that

- (i) z is uncorrelated with ϵ , and
- (ii) z is correlated with ω .

Observe that

$$\begin{aligned} \frac{1}{n} \sum_i (z_i - \bar{z}) \nu_i &= \frac{1}{n} \sum_i (z_i - \bar{z}) \beta_0 + \frac{1}{n} \sum_i (z_i - \bar{z}) \omega_i \beta_1 + \frac{1}{n} \sum_i (z_i - \bar{z}) \epsilon_i \\ &= \frac{1}{n} \sum_i (z_i - \bar{z}) \omega_i \beta_1 + \frac{1}{n} \sum_i (z_i - \bar{z}) \epsilon_i \\ &\rightarrow \frac{1}{n} \sum_i (z_i - \bar{z}) \omega_i \beta_1, \quad \text{as } n \rightarrow \infty, \end{aligned}$$

by the properties of z . This suggests the following estimator of β_1 :

$$\hat{\beta}_{1IV} = \frac{\sum (z_i - \bar{z})\nu_i}{\sum (z_i - \bar{z})\omega_i}, \quad (4.8)$$

and the intercept β_0 is estimated by $\hat{\beta}_{0IV} = \bar{\nu} - \hat{\beta}_{1IV}\bar{\omega}$. The supplementary variable z is called instrumental variable, because it is used merely as an instrument in the estimation of parameters. Estimator (4.8) is called the instrumental variable estimator of β_1 . There is no normality assumption made on any variable. The only assumptions regard the instrumental variable: the properties (i) and (ii) above.

It can be shown (see Fuller 1987, Cheng and Van Ness 1999, Chap. 4), (see Cheng and Van Ness, 1999, Chapter 4) that the instrumental variable estimator (4.8) is consistent and asymptotically Normal with variance $\sigma_\epsilon^2 \sigma_z^2 / (\sigma_{z\omega})^2$, where σ_z^2 is the variance of z and $\sigma_{z\omega}$ is the covariance between z and ω .

Consider, now, the structural ME model

$$\omega_i = \chi_i + \epsilon_i, \quad \nu_i = \beta_0 + \beta_1 \chi_i + \xi_i,$$

which can be written as (4.7) with $\epsilon_i = \xi_i - \beta_1 \epsilon_i$. Clearly, the error ϵ_i is correlated with ω_i . Therefore, if a third variable, z , is available with the properties

(i') z is uncorrelated with (ϵ, ξ) ,

(ii') z is correlated with χ ,

then the instrumental variable estimator of β_1 is exactly (4.8), which is consistent and asymptotically Normal with variance

$$\sigma_\epsilon^2 \sigma_z^2 / (\sigma_{z\chi})^2. \quad (4.9)$$

From (4.9), the new variable z has to be uncorrelated with the errors (ϵ, ξ) , but preferably strongly correlated with χ because the (asymptotic) variance of the instrumental variable estimator ought to be as small as possible. More details about inference via instrumental variables can be founded in Fuller (1987) and Carroll *et al.* (1995).

4.3 ME models as a Missing Data Problem: the Bayesian Approach

The classical measurement error problem discussed until now states that one set of variables, χ , is never observable, i.e. always missing. As such, the ME model is a particular case of missing data problem but with supplementary information about χ in the form of surrogate, ω .

Henceforth, by *surrogate* we mean that ω and ν are conditionally independent given χ .

The earlier measurement error model literature has pursued the functional modeling approach without any distributional specification on χ 's variables. Instead, the most recent missing data literature for ME models, as for example Berry *et al.* (2002), pursues likelihood and Bayesian methods, i.e. follows the structural modeling approach where χ is assumed as a random variables. For these last models, the likelihood definition when the variable χ is unobserved becomes a basic inference tool.

4.3.1 Bayesian methods when χ variable is unobservable

The classical ME model defined in (4.2)-(4.3), is specified if the joint distribution for ν , ω and χ is specified:

$$Pr(\nu, \omega, \chi) = Pr(\nu \mid \omega, \chi)Pr(\omega, \chi). \quad (4.10)$$

When ω is a surrogate, i.e. $\omega \mid \chi$ is independent from $\nu \mid \chi$, it provides no additional information about ν when χ is known, so that, (4.10) is

$$\begin{aligned} Pr(\nu, \omega, \chi) &= Pr(\nu \mid \chi, \alpha)Pr(\omega, \chi) \\ &= Pr(\nu \mid \chi, \alpha)Pr(\omega \mid \chi, \beta)Pr(\chi, \gamma) \end{aligned} \quad (4.11)$$

Equation (4.11) has three components: the first, $Pr(\nu \mid \chi, \alpha)$, is the underlying model of primary interest, with unknown parameter α ; the second, $Pr(\omega \mid \chi, \beta)$, is the error model for ω giving the ‘true’ covariates χ , with unknown parameter β ; finally the third component is the joint distribution of the χ and its parameter γ . Instead, considering the Berkson model, the joint distribution is still (4.10) but the unobserved variable χ is related to ω by $\chi = \omega + \epsilon$. In this case, the likelihood is given by the product over the sample of the following terms:

$$\begin{aligned} Pr(\nu, \omega, \chi) &= Pr(\nu \mid \chi, \alpha)Pr(\omega, \chi) \\ &= Pr(\nu \mid \chi, \alpha)Pr(\chi \mid \omega, \delta)Pr(\omega, \phi) \end{aligned} \quad (4.12)$$

Both (4.11) and (4.12) describe a hierarchical structure. The only difference between (4.11) and (4.12) is the way in which the joint distribution for $Pr(\omega, \chi)$ is factorized. In the Berkson error model (4.12), the unobserved true variable, χ , is conditional to the surrogate, ω besides the unknown parameter δ . In (4.11), instead, the surrogate is observed conditionally on the unobserved true χ , with the consequent need to specify a probability distribution on χ , i.e. it is needed to follow a structural modeling approach.

There are debates in literature regarding the appropriateness of (4.11) or (4.12) in different applications, see e.g. Richardson and Gilks (1993) or Section 7.3 of Carroll *et al.* (1995). However, the usual distinction between structural and functional models, namely whether unknown covariates, χ 's, are random variable or fixed constants, is blurred in the Bayesian framework, where all parameters are random. In fact, the Bayesian distinction between functional and structural models is that under

the latter the χ 's have a common parametric distribution. So that, the structural ME model has $Pr(\chi_i | \gamma)$ i.i.d., that is, of the same form for all sample variable χ_i , with $i = 1, \dots, n$ and with a common parameter γ .

In the functional model, instead, $Pr(\chi_i | \gamma)$ are dependent on the i^{th} observation. In this case, a possible approach can be a hierarchical model, where the $Pr(\chi | \gamma)$ is independent of i and the parameter γ follows its own independent probability distribution.

A second possibility, intermediate between the functional and the structural approaches, is to specify a flexible distribution, as in Mallick and Gelfand (1995), derived following a modified Breckson model specification.

In order to fit the ME models (4.11) and (4.12) as fully Bayesian models, a prior specification on all unknown model parameters is needed. Generally speaking, denoting with $\Lambda = \{\alpha, \beta, \gamma, \delta, \phi\}$ the parameter vector, we can refer to posterior density distribution proportionally as the product of the likelihood and the priors. For the classical ME model (4.11), the posterior density is

$$\begin{aligned} Pr(\Lambda | \nu, \omega, \chi) &\propto Pr(\nu | \chi, \alpha) Pr(\omega | \chi, \beta) Pr(\chi | \gamma) Pr(\alpha, \beta, \gamma) \\ &\propto Pr(\nu | \chi, \alpha) Pr(\omega | \chi, \beta) Pr(\chi | \gamma) Pr(\alpha) Pr(\beta) Pr(\gamma) \end{aligned} \quad (4.13)$$

where the factorization $Pr(\alpha, \beta, \gamma) = Pr(\alpha) Pr(\beta) Pr(\gamma)$ is suggested by the hierarchical structure in (4.11), (see Mallick and Gelfand, 1995).

In a similar way we obtain the posterior density distribution for the Berkson model, given by

$$\begin{aligned} Pr(\Lambda | \nu, \omega, \chi) &\propto Pr(\nu | \chi, \alpha) Pr(\chi | \omega, \delta) Pr(\omega | \phi) Pr(\alpha, \delta, \phi) \\ &\propto Pr(\nu | \chi, \alpha) Pr(\chi | \omega, \delta) Pr(\omega | \phi) Pr(\alpha) Pr(\delta) Pr(\phi). \end{aligned} \quad (4.14)$$

Since the posterior densities are specified, inference is performed using the Markov Chain Monte Carlo algorithms as Gibbs sampling, (Gelfand and Smith, 1990) or Metropolis-Hastings (Smith and Roberts, 1993), to obtain random draws from these posterior distributions.

Further approaches as a semiparametric method for functional models can be used for inference in ME models with missing data, (see e.g. Carroll *et al.*, 1995, Chap. 9).

Henceforth, we deal with ME models by a Bayesian approach applying (4.13) or (4.14).

4.4 Wrapped ME model

As mentioned above, the main properties of the wrapping approach are the flexibility and easy applicability to several probability models and processes. Now we exhibit the application of the wrapping procedure to quite complicated model frameworks as the ME models. The ME models issue discussed until now concerns the general literature on the *in line* measurement error models. Here we show the corresponding *wrapped* version of ME models (WME).

From the ME model definitions in (4.2)-(4.3), inherent the classical ME model, as well as the definition (4.6) for the Berkson model, we note that all components of the measurement error models are additive. This feature is very important because it allows us to apply the wrapping procedure $X_w = Y(mod2\pi)$ described in Section 2.4.3, componentwise.

For this purpose, consider the property (a) of the wrapping distribution in Section 2.4.3 applied to in line variables for deriving the corresponding circular wrapped (w) variables:

$$(y_1 + y_2)_w = y_{1w} + y_{2w}, \quad (4.15)$$

i.e. the wrapping procedure is a homomorphism from \mathbb{R} to the circle. Applying this property to each additive component of the ME model we can obtain the corresponding wrapped version of the ME models. So, given the classical ME model

$$\begin{aligned} \nu &= \gamma + \xi, \quad \text{with } \gamma = \beta_0 + \beta_1\chi \\ \omega &= \chi + \epsilon \end{aligned}$$

the corresponding wrapped classical ME model is

$$\nu_w = (\gamma + \xi)_w = \gamma_w + \xi_w \quad (4.16)$$

$$\omega_w = (\chi + \epsilon)_w = \chi_w + \epsilon_w \quad (4.17)$$

Similarly, the wrapped version of the Berkson ME model is

$$\nu_w = (\gamma + \xi)_w = \gamma_w + \xi_w \quad (4.18)$$

$$\chi_w = (\omega + \epsilon)_w = \omega_w + \epsilon_w \quad (4.19)$$

Thus, following the wrapping approach procedure (see Chapter 3), we can apply to circular variable any in line statistical model. Besides, property (4.15) allows to preserve the ME model structure for the circular variable as well.

The relations between the variables ν , χ , ω assumed in the in line ME model, as for example the linear relationship between ν and χ given by the β 's coefficients, can change for the circular variables ν_w , χ_w in the WME model. But it is worth to remind that the most remarkable advantages of the wrapping approach are the assumption for which the circular variables are thought as coming from their own corresponding in line variables, and the correspondence between in line and circular parameters. So, conditionally to the wrapping coefficients k 's, it is possible to *unwrap* the circular variables to apply the in line statistical model and inferential methods to the circular variables directly (see Section 2.4.3). Then, as said, the corresponding circular parameter estimations are obtained using the correspondence that exists between the in line and circular parameters.

As an example, assume the Normal distribution of the errors, ξ and ϵ , in the classic ME model, so that also the variables ν and ω are Normal distributed and, in particular, $\nu \sim N(\gamma, \sigma_\xi^2)$, where σ_ξ^2 is the variance of ξ . Under this hypothesis, when γ is estimated, i.e. $\gamma = \beta_0 + \beta_1\chi$ is estimated, we can obtain the corresponding circular estimate by $\gamma_w = \gamma(mod2\pi)$. Similarly, we can derive the circular variance applying

the relationship (2.51), $\rho = e^{-\sigma^2/2}$, between the in line and circular variance of the wrapped Normal distribution, (see properties (i) and (ii) of the wrapped Normal distribution in Section 2.4.3). In this sense, all the hypotheses on the errors and on the relationship between variables of the ME model hold in WME model by assuming the corresponding wrapped distribution on circular variables.

Moreover, as mentioned in Section 3.1, the conditioning on k 's parameters leads to choose for a Bayesian parameter estimation approach. For the same reason, also for the WME model, the Bayesian parameter estimation described in Section 4.3 is the most suitable.

Now we check what has just been mentioned about the correspondence between the in line and circular ME model. In particular, we focus on the analysis of the error variances effect in the ME model and in its corresponding WME model. As the side conditions reveal, the error variance has a very important role in parameter estimation and in depiction of the ME model itself. The next section aims to study these aspects.

4.5 Relationship between in line and Wrapped ME Model

We present some simulated examples in order to illustrate the relationship between the ME model and its corresponding wrapped ME model.

For this purpose, let $Y = X + 2K\pi$ be the wrapped Normal model representation, where Y is a in line Normal random variable and X the corresponding circular random variable with support $[0, 2\pi)$. We assume the following ME model

$$Y_{obs} = Y_{true} + \xi_Y \quad (4.20)$$

where Y_{obs} is the observed variable and ξ_Y is a measurement Normal error with zero mean and variance $\sigma_{\xi_Y}^2$. The latent unobserved variable Y_{true} is supposed normally distributed with mean μ_{true} and variance $\sigma_{Y_{true}}^2$.

The corresponding wrapped variables are obtained by the wrapping procedure as follows: $X_{true} = Y_{true} \pmod{2\pi}$ and $X_{obs} = Y_{obs} \pmod{2\pi}$. So, given the wrapped variables X_{true} and X_{obs} , it is interesting to know how $\xi_X = (X_{obs} - X_{true})$ is distributed; i.e. we are interested in investigating if the wrapping procedure has an influence, and how, on the parameter estimation of the ME model.

Moreover, a sensitivity analysis to the error variance on the ME model parameter estimation is carried out through the analysis of the ratio $\sigma_{\xi_Y}^2 / \sigma_{Y_{true}}^2$ in both cases, in line and wrapped variables. This investigation is particularly important in a ME model context for at least two different aspects: a constant ratio $\sigma_{\xi_Y}^2 / \sigma_{Y_{true}}^2$ verifies the side condition (a) and (b) of Section 4.2 about the ME model identifiability; furthermore, the ratio provides a characterization of the ME model itself indicating the incidence of the error on the whole variability of the model.

For this study, we use eleven different simulated data sets obtained drawing both Y_{true} and ξ_Y from a Normal distribution and computing Y_{obs} by their sum. The

wrapped variables are obtained by wrapping the in line variables as usual. For each of the in line and wrapped data sets a graphical representation and a summarizing table of the estimated parameters are reported in Figures 4.1-4.2 and in Tables 4.1-4.2 respectively. Whereas the complete inference results achieved by the parameter estimation procedure showed in Section 3.1 are reported in Table 4.3.

In Table 4.1 we report six simulated in line and corresponding wrapped data sets in which we analyze the effect on the estimation of the $\sigma_{Y_{true}}^2$ (data sets A, B, C) and the effect of the σ_{ξ}^2 (data sets D, E, F) separately.

In the first three simulated data sets the variance of Y_{true} is fixed equal to 1 whereas the variance σ_{ξ}^2 changes its value giving as many variance ratios equal to 0.25, 1 and 2.25 respectively. The same ratios are accomplished fixing the σ_{ξ}^2 equal to 0.25 and varying the variance $\sigma_{Y_{true}}^2$ in the last three examples.

In general, we can say that a correspondence occurs between the Gaussian parameters and the wrapped Normal parameters and this correspondence persists in a ME model context as well. To stress this issue, we report, in the right column of Tables 4.1-4.2 inherent to wrapped Normal distribution, the circular parameter, ρ , and the corresponding in line parameter σ^2 , the last one computed by transformation (2.51): $\rho = e^{-\sigma^2/2}$. Moreover, we underline the sensitivity of wrapped parameter estimation to data variability: when the variability increases less correspondence occurs between in line and wrapped parameters. This issue arises independently from the ratio value. As a matter of fact, analyzing estimates in Table 4.3 we realize that the posterior interval size increases more according to the variables variability than to the variance ratio value. As an example, from Table 4.3, we can see that the posterior interval sizes of the parameter estimates for data set *C* (0.74 and 2.00 for $\sigma_{X_{true}}^2$ and $\sigma_{\xi_X}^2$ respectively) are larger than the posterior interval sizes for the parameters in data set *F* (equal to 0.14 and 0.11) even though the variances ratio, equal to 2.25, is the same for both data sets.

This aspect is even more evident for the data sets *G-K* where the ratio is constantly equal to 1/3 but the posterior interval sizes increase as the scale value of the variance increases, i.e. the estimates become less reliable as the variability of the data increases. This is perfectly coherent with the simulation-based outcome achieved in Section 3.3 about the variability effect on the parameter estimate accuracy.

4.6 Discussion

In this chapter we introduce general definition and aspects of the measurement error models and analyze the main inferential approach and methods for their estimation. An interesting aspect of the ME model as missing data approach is investigated, leading to some important considerations about the Bayesian inferential approach. Moreover, the full Bayesian model for ME model is specified.

After that the classical in line ME model is presented, the corresponding wrapped ME model is introduced. By using one of the properties of the wrapping distributions, we are able to derive the wrapped ME model such that the structure and the main properties of the in line ME model are conserved. Finally, the relationship

Case	Normal Distribution $Y \sim N(\tilde{\mu}, \sigma^2)$	Wrapped Normal Distribution $X \sim WN(\mu = \tilde{\mu}(\text{mod}2\pi), \rho)$
A	$Y_{true} \sim N(0, 1.00)$ $\xi_Y \sim N(0, 0.25)$ $Y_{obs} = (Y_{true} + \xi_Y) \sim N(0, 1.25)$ $\sigma_{\xi_Y}^2 / \sigma_{Y_{true}}^2 = 0.25$	$X_{true} \sim WN(\mu = -0.009, \rho = 0.57 \equiv \sigma^2 = 1.12)$ $\xi_X \sim WN(\mu = 0.044, \rho = 0.88 \equiv \sigma^2 = 0.26)$ $X_{obs} \sim WN(\mu = 0.034, \rho = 0.52 \equiv \sigma^2 = 1.31)$ $\sigma_{\xi_X}^2 / \sigma_{X_{true}}^2 = 0.23$
B	$Y_{true} \sim N(0, 1.00)$ $\xi_Y \sim N(0, 1.00)$ $Y_{obs} \sim N(0, 2.00)$ $\sigma_{\xi_Y}^2 / \sigma_{Y_{true}}^2 = 1$	$X_{true} \sim WN(\mu = 0.009, \rho = 0.58 \equiv \sigma^2 = 1.09)$ $\xi_X \sim WN(\mu = 0.031, \rho = 0.62 \equiv \sigma^2 = 0.95)$ $X_{obs} \sim WN(\mu = -0.021, \rho = 0.30 \equiv \sigma^2 = 2.41)$ $\sigma_{\xi_X}^2 / \sigma_{X_{true}}^2 = 0.87$
C	$Y_{true} \sim N(0, 1.00)$ $\xi_Y \sim N(0, 2.25)$ $Y_{obs} \sim N(0, 3.25)$ $\sigma_{\xi_Y}^2 / \sigma_{Y_{true}}^2 = 2.25$	$X_{true} \sim WN(\mu = 0.008, \rho = 0.59 \equiv \sigma^2 = 1.08)$ $\xi_X \sim WN(\mu = -0.07, \rho = 0.38 \equiv \sigma^2 = 1.95)$ $X_{obs} \sim WN(\mu = -0.021, \rho = 0.31 \equiv \sigma^2 = 2.65)$ $\sigma_{\xi_X}^2 / \sigma_{X_{true}}^2 = 1.81$
D	$Y_{true} \sim N(0, 0.50)$ $\xi_Y \sim N(0, 0.25)$ $Y_{obs} \sim N(0, 1.25)$ $\sigma_{\xi_Y}^2 / \sigma_{Y_{true}}^2 = 0.50$	$X_{true} \sim WN(\mu = -0.008, \rho = 0.77 \equiv \sigma^2 = 0.51)$ $\xi_X \sim WN(\mu = 0.044, \rho = 0.88 \equiv \sigma^2 = 0.26)$ $X_{obs} \sim WN(\mu = 0.034, \rho = 0.52 \equiv \sigma^2 = 1.31)$ $\sigma_{\xi_X}^2 / \sigma_{X_{true}}^2 = 0.51$
E	$Y_{true} \sim N(0, 0.25)$ $\xi_Y \sim N(0, 0.25)$ $Y_{obs} \sim N(0, 0.50)$ $\sigma_{\xi_Y}^2 / \sigma_{Y_{true}}^2 = 1$	$X_{true} \sim WN(\mu = -0.11, \rho = 0.88 \equiv \sigma^2 = 0.26)$ $\xi_X \sim WN(\mu = 0.044, \rho = 0.88 \equiv \sigma^2 = 0.26)$ $X_{obs} \sim WN(\mu = -0.09, \rho = 0.79 \equiv \sigma^2 = 0.47)$ $\sigma_{\xi_X}^2 / \sigma_{X_{true}}^2 = 1$
F	$Y_{true} \sim N(0, 0.11)$ $\xi_Y \sim N(0, 0.25)$ $Y_{obs} \sim N(0, 0.36)$ $\sigma_{\xi_Y}^2 / \sigma_{Y_{true}}^2 = 2.25$	$X_{true} \sim WN(\mu = -0.009, \rho = 0.95 \equiv \sigma^2 = 0.11)$ $\xi_X \sim WN(\mu = 0.044, \rho = 0.88 \equiv \sigma^2 = 0.26)$ $X_{obs} \sim WN(\mu = -0.09, \rho = 0.84 \equiv \sigma^2 = 0.34)$ $\sigma_{\xi_X}^2 / \sigma_{X_{true}}^2 = 2.36$

Table 4.1: Table of simulated in line distributions and corresponding wrapped distributions. In the A, B and C cases the unobserved variable variance is fixed while in the D, E and F cases the fixed parameter is the error variance.

Case	Normal Distribution $Y \sim N(\tilde{\mu}, \sigma^2)$	Wrapped Normal Distribution $X \sim WN(\mu = \tilde{\mu}(\text{mod}2\pi), \rho)$
G	$Y_{true} \sim N(0, 0.03)$ $\xi_Y \sim N(0, 0.01)$ $Y_{obs} = (Y_{true} + \xi_Y) \sim N(0, 0.04)$ $\sigma_{\xi_Y}^2 / \sigma_{Y_{true}}^2 = 0.33$	$X_{true} \sim WN(\mu = 0.02, \rho = 0.98 \equiv \sigma^2 = 0.03)$ $\xi_X \sim WN(\mu = -0.006, \rho = 0.99 \equiv \sigma^2 = 0.01)$ $X_{obs} \sim WN(\mu = 0.04, \rho = 0.98 \equiv \sigma^2 = 0.04)$ $\sigma_{\xi_X}^2 / \sigma_{X_{true}}^2 = 0.33$
H	$Y_{true} \sim N(0, 0.15)$ $\xi_Y \sim N(0, 0.05)$ $Y_{obs} \sim N(0, 0.20)$ $\sigma_{\xi_Y}^2 / \sigma_{Y_{true}}^2 = 0.33$	$X_{true} \sim WN(\mu = -0.003, \rho = 0.93 \equiv \sigma^2 = 0.14)$ $\xi_X \sim WN(\mu = -0.002, \rho = 0.97 \equiv \sigma^2 = 0.05)$ $X_{obs} \sim WN(\mu = -0.08, \rho = 0.91 \equiv \sigma^2 = 0.18)$ $\sigma_{\xi_X}^2 / \sigma_{X_{true}}^2 = 0.35$
I	$Y_{true} \sim N(0, 0.30)$ $\xi_Y \sim N(0, 0.10)$ $Y_{obs} \sim N(0, 0.40)$ $\sigma_{\xi_Y}^2 / \sigma_{Y_{true}}^2 = 0.33$	$X_{true} \sim WN(\mu = 0.005, \rho = 0.86 \equiv \sigma^2 = 0.30)$ $\xi_X \sim WN(\mu = 0.01, \rho = 0.96 \equiv \sigma^2 = 0.09)$ $X_{obs} \sim WN(\mu = -0.06, \rho = 0.84 \equiv \sigma^2 = 0.34)$ $\sigma_{\xi_X}^2 / \sigma_{X_{true}}^2 = 0.30$
J	$Y_{true} \sim N(0, 0.90)$ $\xi_Y \sim N(0, 0.30)$ $Y_{obs} \sim N(0, 1.20)$ $\sigma_{\xi_Y}^2 / \sigma_{Y_{true}}^2 = 0.33$	$X_{true} \sim WN(\mu = 0.07, \rho = 0.66 \equiv \sigma^2 = 0.83)$ $\xi_X \sim WN(\mu = 0.11, \rho = 0.95 \equiv \sigma^2 = 0.32)$ $X_{obs} \sim WN(\mu = 0.09, \rho = 0.60 \equiv \sigma^2 = 1.01)$ $\sigma_{\xi_X}^2 / \sigma_{X_{true}}^2 = 0.38$
K	$Y_{true} \sim N(0, 1.50)$ $\xi_Y \sim N(0, 0.50)$ $Y_{obs} \sim N(0, 2.00)$ $\sigma_{\xi_Y}^2 / \sigma_{Y_{true}}^2 = 0.33$	$X_{true} \sim WN(\mu = -0.08, \rho = 0.43 \equiv \sigma^2 = 1.68)$ $\xi_X \sim WN(\mu = -0.04, \rho = 0.79 \equiv \sigma^2 = 0.46)$ $X_{obs} \sim WN(\mu = 0.13, \rho = 0.28 \equiv \sigma^2 = 2.54)$ $\sigma_{\xi_X}^2 / \sigma_{X_{true}}^2 = 0.27$

Table 4.2: Table of simulated in line distributions and corresponding wrapped distributions. The variance ratio $\sigma_{\xi}^2 / \sigma_{Y_{true}}^2$ is constant.

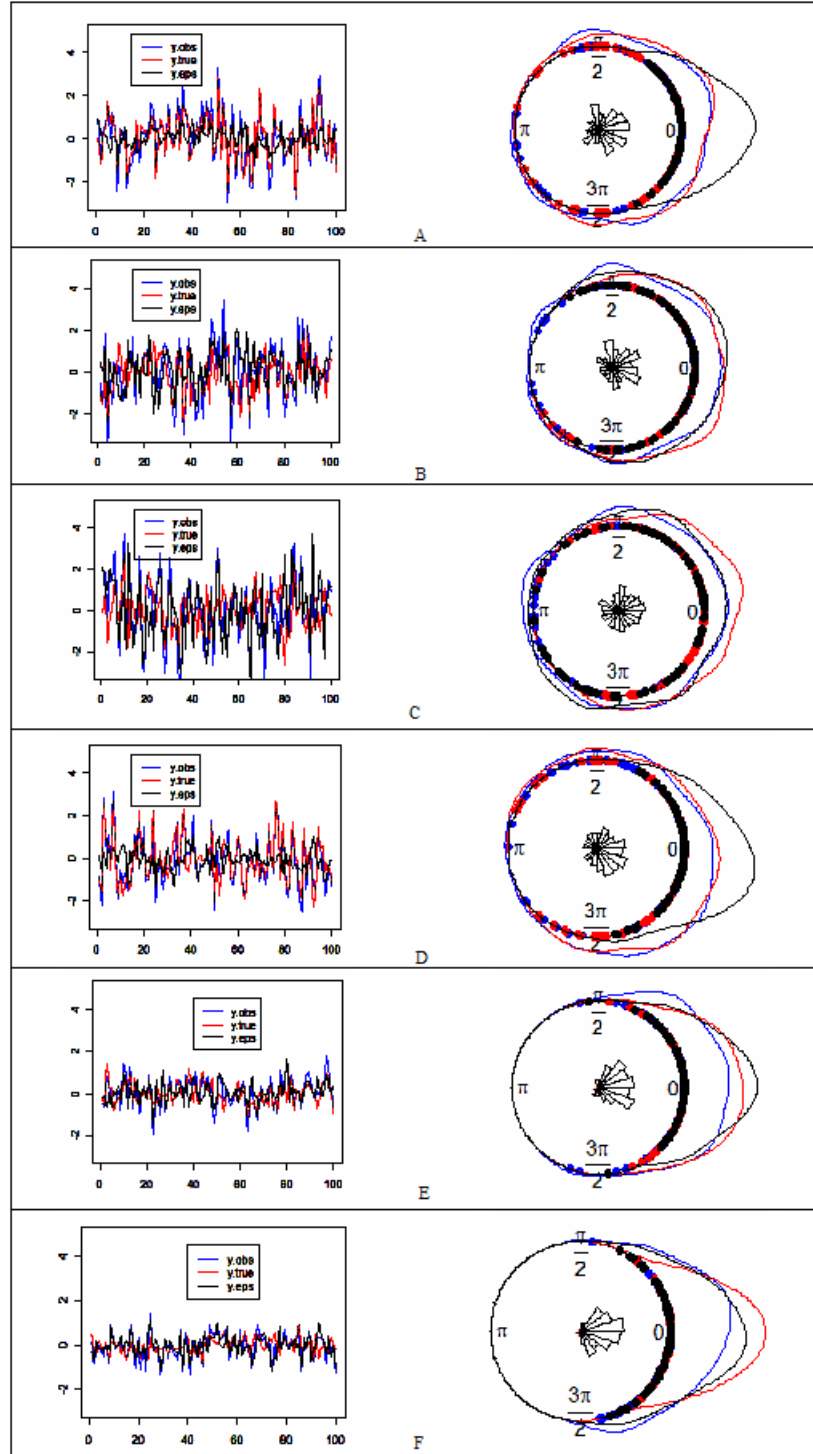


Figure 4.1: Plot of simulated in line distributions (left plots) and corresponding wrapped distributions (right rose diagram plots) inherent to distributions reported in Table 4.1

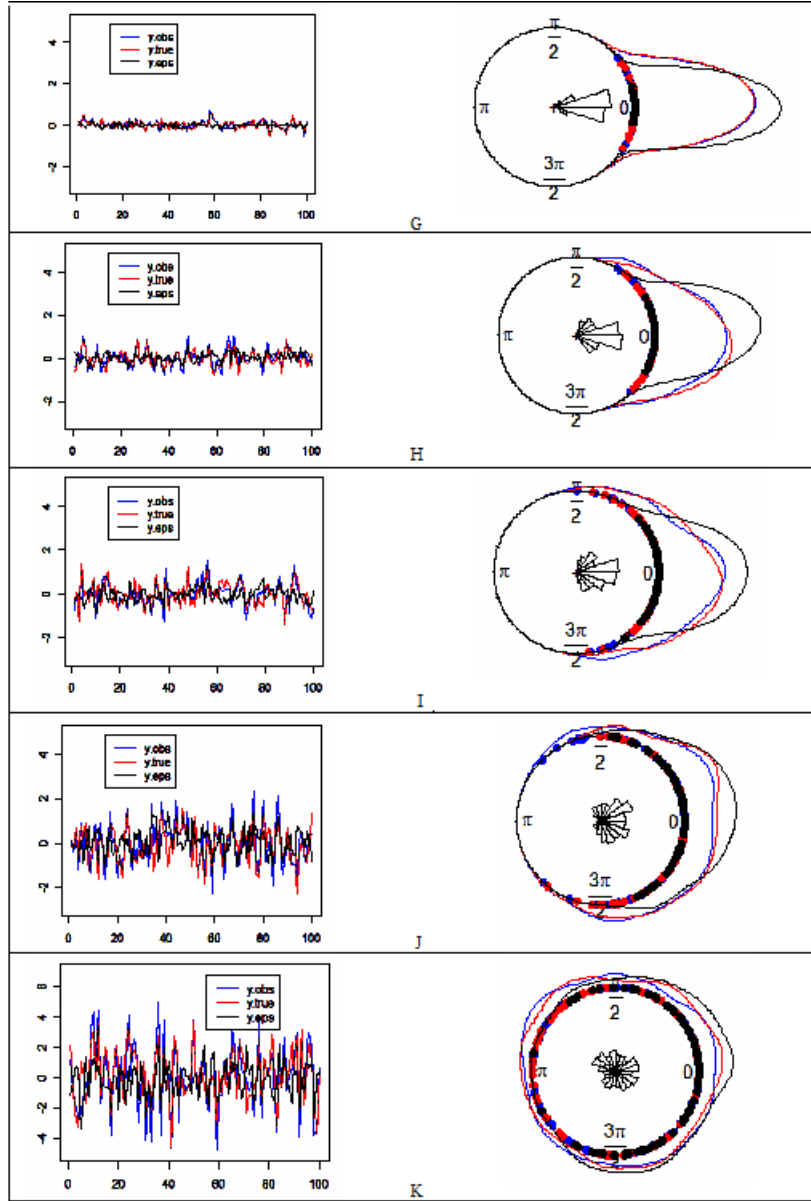


Figure 4.2: Plot of simulated in line distributions (left plots) and corresponding wrapped distributions (right rose diagram plots) inherent to distributions reported in Table 4.2

Data Set	Ratio $\sigma_\xi^2/\sigma_{true}^2$	Latent Variable $_{true}$			Error Variable ξ		
		In line	Wrapped	95%	In line	Wrapped	95%
		True value $\sigma_{Y_{true}}^2$	Estimate $\sigma_{X_{true}}^2$	Posterior Median	True value $\sigma_{\xi_Y}^2$	Estimate $\sigma_{\xi_X}^2$	Posterior Median
<i>A</i>	0.25	1.00	1.12	[0.77, 1.47]	0.25	0.26	[0.20, 0.33]
<i>B</i>	1.00	1.00	1.09	[0.71, 1.49]	1.00	0.95	[0.57, 1.46]
<i>C</i>	2.25	1.00	1.08	[0.72, 1.46]	2.25	1.95	[0.87, 2.87]
<i>D</i>	0.25	1.00	1.10	[0.69, 1.51]	0.25	0.24	[0.18, 0.33]
<i>E</i>	1.00	0.25	0.26	[0.21, 0.34]	0.25	0.26	[0.20, 0.34]
<i>F</i>	2.25	0.11	0.11	[0.04, 0.18]	0.25	0.26	[0.21, 0.32]
<i>G</i>	0.33	0.03	0.03	[0.01, 0.05]	0.01	0.01	[0.007, 0.013]
<i>H</i>	0.33	0.15	0.14	[0.05, 0.25]	0.05	0.05	[0.02, 0.08]
<i>I</i>	0.33	0.30	0.30	[0.24, 0.37]	0.10	0.09	[0.03, 0.17]
<i>J</i>	0.33	0.90	0.83	[0.56, 1.24]	0.30	0.32	[0.22, 0.41]
<i>K</i>	0.33	1.50	1.68	[0.98, 2.36]	0.50	0.46	[0.26, 0.78]

Table 4.3: ME model variance estimates

between in line and its corresponding wrapped ME model is investigated with a particular attention to the sensitivity analysis to the error variance on parameter estimation. This last analysis is carried out by simulation-based study, and it confirms the evidence came out in Section 3.3 about the variability effect on parameter estimation: even in the case of ME model, the data variability of the circular data influences the estimation reliability.

Chapter 5

Spatial Modeling

In this chapter we present the essential elements of the standard *in line* spatial model and the corresponding wrapped spatial model. In particular we provide a first overview on the spatial data and models, then we focus on point-referenced data describing the main elements of this of model. Successively we introduce the wrapped spatial model showing some applications to simulated and real data.

5.1 Introduction to Spatial Models

Spatial data can be thought of as a realization of a random process (or random field)

$$\{\mathbf{Y}(\mathbf{s}) : \mathbf{s} \in D\} \quad (5.1)$$

where \mathbf{s} is a vector of locations and D is a subset of \mathbb{R}^d .

The definition of a spatial data set induces a classification of the data in three main categories according to the definition of D or \mathbf{s} :

- a) point-referenced data;
- b) areal data;
- c) point pattern data.

Case a) is often referred to as *geostatistical* (Matheron, 1962) data where \mathbf{s} varies *continuously* over D .

Also the second type of spatial data is often denoted with another name: lattice data. A lattice of locations evokes the idea of regularly spaced points in \mathbb{R}^d linked to the neighbors. For this reason in areal (or lattice) data, D (it can be of regular or irregular shape) is partitioned into a finite number of areal units with well-defined boundaries. Thus, the locations $\mathbf{s} \in D$, in this case, are the regions (or *blocks*) themselves. In this context, data are typically sums or averages of variable over blocks. To introduce the spatial association, it is necessary in this case to define a neighborhood structure based on the arrangement of the blocks in the map. Once the neighborhood is defined, models resembling autoregressive time series models are considered. The most important models that incorporate such neighborhood

information are the *simultaneously autoregressive (SAR)* and *conditionally autoregressive (CAR)* models, introduced by Whittle (1954) and Besag (1974) respectively. The third case c) arises when the response variable is the location of “events”. In point-pattern data, in fact, the variable to be analyzed is often fixed and the locations \mathbf{s} are thought as random, i.e., D itself is random. Such data are often of interest in studies of event clustering, where the goal is to determine whether an observed spatial point pattern is an example of a cluster process or merely the result of a random event process over space.

5.2 Point-referenced Data Models

As mentioned above, the basic concept underlying the spatial theory is to think of data as resulting from observations of the stochastic process $Y(\mathbf{s})$, defined in (5.1). In point-referenced data, D is a fixed subset of the d -dimensional Euclidean space. In situations where $d > 1$, the process is often referred to as *spatial process*. To introduce the statistical model for spatial data, it is necessary to give some basic definition and concepts as stationarity, isotropy and variograms.

5.2.1 Stationarity

The random process (5.1) is usually defined through the finite-dimensional distribution

$$F_{s_1, \dots, s_n}(y_1, \dots, y_n) = Pr(Y(s_1) < y_1, \dots, Y(s_n) < y_n), \quad n \geq 1 \quad (5.2)$$

which must satisfy Kolmogorov’s conditions of symmetry (i.e., the probability distribution F remains invariant when y_i and s_i are subjected to the same permutation) and consistency (i.e., $F_{s_1, \dots, s_{k+1}}(y_1, \dots, y_k, \infty, \dots, \infty) = F_{s_1, \dots, s_k}(y_1, \dots, y_k)$). Considering that data represent sampling of a *single* realization, inference on \mathbf{Y} without further assumptions is unfeasible.

One strongest assumption regards stationarity. The process is said to be *strictly stationary* if, for any given $n > 1$, any set of n sites $\{s_1, \dots, s_n\}$ and any $\mathbf{h} \in \mathbb{R}^d$, the distribution of $(Y(s_1), \dots, Y(s_n))$ is the same as that of $(Y(s_1 + \mathbf{h}), \dots, Y(s_n + \mathbf{h}))$. The strictly stationarity legitimizes, the use of the observations $y(s_i)$ as different realizations of the same random variable. Nevertheless, this condition is rarely satisfied, as a consequence, a less restrictive condition is *weak stationarity* (also called second-order stationarity). Cressie (1993, pag.53) defines a spatial process to be weakly stationary if:

$$\mu(\mathbf{s}) \equiv \mu(\text{i.e., the process has a constant mean})$$

and

$$Cov(Y(s), Y(s + \mathbf{h})) = C(\mathbf{h})$$

for all $\mathbf{h} \in \mathbb{R}^d$ such that \mathbf{s} and $\mathbf{s} + \mathbf{h}$ both lie within D .

Weak stationarity implies that the covariance relationship between the values of the process at any two locations can be summarized by a covariance function $C(\mathbf{h})$,

and this function depends only on the distance vector \mathbf{h} . Note that assumed the existence of all variances, strong stationarity implies weak stationarity. The converse is not true in general but it holds for the Gaussian process. Defining, in fact, as *Gaussian* the process $Y(\mathbf{s})$ that has, for any $n > 1$ and for any set of sites $\{s_1, \dots, s_n\}$, $\mathbf{Y} = (Y(s_1), \dots, Y(s_n))^T$ distributed as a multivariate Normal, the weak stationarity, i.e., the stationarity on the first two moments, implies strictly stationarity as well.

There is another type of stationarity that leads to definition of the another crucial instrument for geostatistics: the variogram.

5.2.2 Variogram

The *intrinsic stationarity* is defined through the first difference:

$$\begin{aligned} E(Y(\mathbf{s} + \mathbf{h}) - Y(\mathbf{s})) &= 0, \\ E(Y(\mathbf{s} + \mathbf{h}) - Y(\mathbf{s}))^2 &= Var(Y(\mathbf{s} + \mathbf{h}) - Y(\mathbf{s})) = 2\gamma(\mathbf{h}) \end{aligned} \quad (5.3)$$

Equation 5.3 makes sense only if the left-hand side depends only on \mathbf{h} (so that the right-hand side can be written at all), and not the particular choice of \mathbf{s} . In this case, then, the process $Y(\mathbf{s})$ is intrinsically stationary.

Function $2\gamma(\mathbf{h})$ is the *variogram* (Matheron, 1962), whereas $\gamma(\mathbf{h})$ is called the *semivariogram*. From definition (5.3) it easy to verify some properties of the variogram:

- (i) $\gamma(-\mathbf{h}) = \gamma(\mathbf{h})$ i.e., the variogram is a even function;
- (ii) $\gamma(\mathbf{0}) = 0$ i.e., the variogram is continuous at the origin.

The first property is always verified, while the second one is not granted. It is possible that as $\mathbf{h} \rightarrow \mathbf{0}$, $\gamma(\mathbf{h}) \rightarrow \tau^2 > 0$. Matheron (1962) defined this quantity the *nugget effect* specifying that the only possible reason for $\tau^2 > 0$ is the measurement error. This leads to add a white-noise process to the process with continuous sample path.

Another important property of the variogram is its relationship with the covariance function $C(\mathbf{h})$:

$$\begin{aligned} 2\gamma(\mathbf{h}) &= Var(Y(\mathbf{s} + \mathbf{h}) - Y(\mathbf{s})) \\ &= Var(Y(\mathbf{s} + \mathbf{h})) + Var(Y(\mathbf{s})) - 2Cov(Y(\mathbf{s} + \mathbf{h}), Y(\mathbf{s})) \\ &= C(\mathbf{0}) + C(\mathbf{0}) - 2C(\mathbf{h}) \\ &= 2(C(\mathbf{0}) - C(\mathbf{h})) \end{aligned}$$

Thus

$$\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h}) \quad (5.4)$$

where $C(\mathbf{0}) = \sigma^2$ is the variance of the process, also called *partial sill*. The *sill* is properly defined by $\lim_{\|\mathbf{h}\| \rightarrow \infty} \gamma(\|\mathbf{h}\|) = \tau^2 + \sigma^2$. Finally, the last parameter that characterizes the variogram, together with the sill and the nugget, is the *range* that

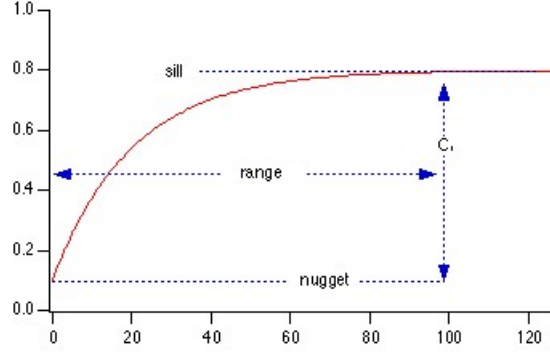


Figure 5.1: Variogram and its parameters.

is the value of $\|\mathbf{h}\|$ at which $\gamma(\|\mathbf{h}\|)$ reaches its ultimate level (the sill). A graphical representation of these parameters is depicted in Figure 5.1.

A variogram to be valid must satisfy a negative definiteness condition (Matheron, 1971b) given by

$$\sum_i \sum_j a_i a_j \gamma(s_i - s_j) \leq 0, \quad (5.5)$$

for any finite number of spatial locations s_1, \dots, s_n and for any set of constants a_1, \dots, a_n such that $\sum_i a_i = 0$.

Other conditions for a valid variogram are discussed in Cressie (1993).

5.2.3 Isotropy

A third important concept is the isotropy. If the semivariogram function $\gamma(\mathbf{h})$ depends upon the distance vector only through its length $\|\mathbf{h}\|$, then the process is *isotropic*, if not, it is *anisotropic*. Thus, for an isotropic process, $\gamma(\mathbf{h})$ is a real-valued function of a univariate argument and can be written as $\gamma(\|\mathbf{h}\|)$.

Isotropic processes are popular because for them a number of relatively simple parametric forms are available as candidates for the semivariogram. Denoting $\|\mathbf{h}\|$ by t , we now consider some of the more important forms.

Linear:

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 t & \text{if } t > 0, \tau^2 > 0, \sigma^2 > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Note that $\gamma(t) \rightarrow \infty$ as $t \rightarrow \infty$, and so this semivariogram does not correspond to a weakly stationary process, but only to an intrinsically stationary process.

Spherical:

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 t & \text{if } t \geq 1/\phi, \\ \tau^2 + \sigma^2 \left\{ \frac{3\phi t}{2} - \frac{1}{2}(\phi t)^3 \right\} & \text{if } 0 < t \leq 1/\phi, \\ 0 & \text{otherwise.} \end{cases}$$

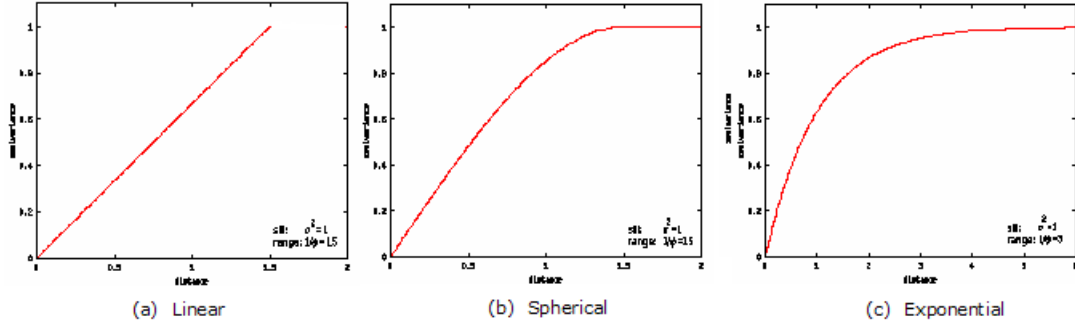


Figure 5.2: Theoretical semivariogram: (a) linear, (b) spherical, (c) exponential.

The value $t = 1/\phi$ is the range while ϕ is often referred to as the *decay* parameter. The spherical semivariogram is valid in $d = 1, 2$ or 3 . The spherical form does give rise to a stationary process and so the corresponding covariance function is easily computed.

Exponential:

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2(1 - \exp(-\phi t)) & \text{if } t > 0, \\ 0 & \text{otherwise.} \end{cases}$$

The exponential has an advantage over the spherical in that it is simpler in functional form and it is valid in all dimensions.

Note from Figure 5.2 (c) that the sill is reached only asymptotically, meaning that the range $1/\phi$ is infinite. In cases like this, it is appropriate to introduce the notion of *effective range*, that is the distance at which there is essentially no lingering spatial correlation. A common definition of effective range \tilde{t} , is the distance at which this correlation drops to only 0.05. Setting $\exp(-\phi\tilde{t})$ equal to this value we obtain $\tilde{t} \approx 3/\phi$, since $\log(0.05) \approx -3$. The procedure to find a distance over that the covariance function is supposed valued under a fix value is commonly used to decide reasonable value for the range parameter.

Gaussian:

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2(1 - \exp(-\phi^2 t^2)) & \text{if } t > 0, \\ 0 & \text{otherwise.} \end{cases}$$

The Gaussian variogram yields very smooth realizations of the spatial process.

Wave:

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2(1 - \frac{\sin(\phi t)}{\phi t}) & \text{if } t > 0, \\ 0 & \text{otherwise.} \end{cases}$$

The Wave variogram is as example of not monotonic increase due to *sin* function.

Model	Variogram $\gamma(t)$
Linear	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 t & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases}$
Spherical	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 & \text{if } t \geq 1/\phi \\ \tau^2 + \sigma^2 [\frac{3}{2}\phi t - \frac{1}{2}(\phi t)^3] & \text{if } 0 < t < 1/\phi \\ 0 & \text{otherwise} \end{cases}$
Exponential	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2(1 - \exp(-\phi t)) & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases}$
Gaussian	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2(1 - \exp(-\phi^2 t^2)) & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases}$
Power	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 t^\lambda & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases}$
Wave	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2(1 - \frac{\sin(\phi t)}{\phi t}) & \text{if } t > 0, \\ 0 & \text{otherwise} \end{cases}$
Matern	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 \left[1 - \frac{(2\sqrt{\nu}t\phi)^\nu}{2^{\nu-1}\Gamma(\nu)} K_\nu(2\sqrt{\nu}t\phi)\right] & \text{if } t > 0, \\ 0 & \text{otherwise} \end{cases}$

Table 5.1: Semivariograms for common parametric isotropic models.

Power law:

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 t^\lambda & \text{if } t > 0, \\ 0 & \text{otherwise.} \end{cases}$$

This semivariogram generalizes the linear one and produces valid intrinsic stationary semivariograms with $0 \leq \lambda < 2$.

Matérn:

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 \left[1 - \frac{(2\sqrt{\nu}t\phi)^\nu}{2^{\nu-1}\Gamma(\nu)} K_\nu(2\sqrt{\nu}t\phi)\right] & \text{if } t > 0, \\ 0 & \text{otherwise,} \end{cases}$$

where K_ν is the modified Bessel function of order ν . This class of semivariograms was originally suggested by Matérn (1986). The parameter $\nu > 0$ controls the smoothness of the realized random field, while ϕ is a spatial scale parameter. We refer to Matérn semivariogram as a class because, with opportune parametrization of ν , we obtain other semivariogram forms as for example the exponential (with $\nu = 1/2$) and the Gaussian (with $\nu \rightarrow \infty$).

The semivariograms above described and their covariance functions are summarized in Tables 5.1 and 5.2 respectively.

5.2.4 Variogram model fitting

Several variogram estimators exist in literature (a comprehensive account of these estimators can be found in Cressie (1993, sec. 2.5)) and perhaps the most customary is the method-of-moment estimator due to Matheron (1962):

$$\hat{\gamma}(t) = \frac{1}{2N(t)} \sum_{(\mathbf{s}_i, \mathbf{s}_j) \in N(t)} [Y(\mathbf{s}_i) - Y(\mathbf{s}_j)]^2, \quad (5.6)$$

Model	Covariance function $C(t)$
Linear	$C(t)$ does not exist
Spherical	$C(t) = \begin{cases} 0 & \text{if } t \geq 1/\phi \\ \sigma^2 [1 - \frac{3}{2}\phi t - \frac{1}{2}(\phi t)^3] & \text{if } 0 < t \leq 1/\phi \\ \tau^2 + \sigma^2 & \text{otherwise} \end{cases}$
Exponential	$C(t) = \begin{cases} \sigma^2 \exp(-\phi t) & \text{if } t > 0 \\ \tau^2 + \sigma^2 & \text{otherwise} \end{cases}$
Gaussian	$C(t) = \begin{cases} \sigma^2 \exp(-\phi^2 t^2) & \text{if } t > 0 \\ \tau^2 + \sigma^2 & \text{otherwise} \end{cases}$
Power	$C(t)$ does not exist
Wave	$C(t) = \begin{cases} \sigma^2 \frac{\sin(\phi t)}{\phi t} & \text{if } t > 0, \\ \tau^2 + \sigma^2 & \text{otherwise} \end{cases}$
Matern	$C(t) = \begin{cases} \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} (2\sqrt{\nu}t\phi)^\nu K_\nu(2\sqrt{\nu}t\phi) & \text{if } t > 0, \\ \tau^2 + \sigma^2 & \text{otherwise} \end{cases}$

Table 5.2: Covariance functions for common parametric isotropic models.

where $N(t)$ is the set of pairs of the points such that $\|\mathbf{s}_i - \mathbf{s}_j\| = t$, and $N(t)$ is the number of the pairs in this set. Although easy and of immediate intuition, this estimator is sensible to outliers. To overcome this drawback, Hawkins and Cressie (1984) proposed a robust variogram estimator.

However, these estimators are not necessarily conditionally negative-definite, so we need to have recourse to valid variograms. For this purpose, the usual approach to fit the variogram consists of plotting the empirical variogram and then choose the theoretical variogram model that best fits the “data”. More formally we can treat this as an estimation problem and estimate the parameter sill, range and nugget following some goodness-of-fit criteria, (see Smith, 2001).

Finally, if the distribution model is available, it is possible to obtain maximum likelihood estimates for the variogram parameters, (see Cressie, 1993, Sec. 2.6). Hierarchical Bayesian approach is similar to likelihood approach although it is easier work directly with the covariance function rather than changing to a partial likelihood in order to introduce the semivariogram.

5.3 Hierarchical Modeling for Univariate Point-referenced Data

From a Bayesian point of view, there are basically two ways to introduce spatial models: either in first-stage specification, to directly model data in a spatial fashion, or in a second-stage specification, to model the spatial structure via the random effects. In the last case, the spatial process is viewed as latent component and data, modeled at the first stage, help to learn about the process. In the last years, a fairly extensive literature on spatial prediction has been developed from a Bayesian perspective, like Ecker and Gelfand (1997), Diggle *et al.* (1998), Karson *et al.* (1999).

5.3.1 Spatial model definition

The basic spatial model is:

$$Y(\mathbf{s}) = \mu(\mathbf{s}) + w(\mathbf{s}) + \xi(\mathbf{s}) \quad \mathbf{s} = (s_1, \dots, s_d) \quad (5.7)$$

where the mean structure, $\mu(\mathbf{s})$, can be expressed through the covariates \mathbf{x} , i.e., $\mu(\mathbf{s}) = \mathbf{x}^T(\mathbf{s})\beta$. The residual is partitioned into two components, a spatial process, $w(\mathbf{s})$, and a non spatial term $\xi(\mathbf{s})$. The first are assumed to be realizations from a stationary Gaussian spatial process with zero mean vector, while $\xi(\mathbf{s})$ are uncorrelated pure error terms. Substantially, $w(\mathbf{s})$ introduce the partial sill (σ^2) and range (ϕ) parameters, while $\xi(\mathbf{s})$ add the nugget effect (τ^2).

5.3.2 Hierarchical Bayesian methods

Consider a general Gaussian linear model $\mathbf{Y} \sim N_n(X\beta, \Sigma)$, where the response variable is $n \times 1$ data vector with known variance and covariance matrix Σ and X is an $n \times p$ matrix of covariates. From a Bayesian point of view, the likelihood and the prior for the unknown parameter vector β are given respectively by:

$$\mathbf{Y} \mid \beta \sim N_d(X\beta, \Sigma),$$

$$\beta \sim N_p(A\alpha, V).$$

Then, it can be shown that the marginal distribution of \mathbf{Y} is

$$\mathbf{Y} \sim N_d(XA\alpha, \Sigma + XVX^T),$$

and the posterior distribution of β is

$$\beta \mid \mathbf{Y} \sim N_p(D\mathbf{d}, D)$$

where $D^{-1} = X^T\Sigma^{-1}X + V^{-1}$ and $\mathbf{d} = X^T\Sigma^{-1}\mathbf{Y} + V^{-1}A\alpha$. Thus, the Bayesian point estimate for parameter β can be found, for example, by averaging of the posterior distribution: $E(\beta \mid \mathbf{Y}) = D\mathbf{d}$.

In a spatial context, where we suppose to collect the data in $\mathbf{Y} = (Y(s_1), \dots, Y(s_d))^T$, we can adopt the same approach upon an appropriate definition of the Σ matrix. For example, in case of a spatial model with nugget effect, the matrix is

$$\Sigma = \sigma^2 H(\phi) + \tau^2 I, \quad (5.8)$$

where H is a correlation matrix with $H_{ij} = \rho(\mathbf{s}_i - \mathbf{s}_j; \phi)$ and ρ is a valid isotropic correlation function on \mathbb{R}^2 indexed by the parameter ϕ .

Let $\Psi = (\beta, \sigma^2, \tau^2, \phi)^T$, a Bayesian solution requires an appropriate prior distribution $Pr(\Psi)$, so that the parameter estimate can be obtained from the posterior distribution notoriously proportional to the product of likelihood and prior:

$$Pr(\Psi \mid \mathbf{Y}) \propto f(\mathbf{Y} \mid \Psi) Pr(\Psi).$$

Then, the full specification of the spatial model comes from the likelihood (as the product of the probability distribution function) and the priors (which are typically chosen independent each other for the different parameters) as:

$$\begin{aligned} \mathbf{Y} \mid \boldsymbol{\Psi} &\sim N(X\boldsymbol{\beta}, \sigma^2 H(\phi) + \tau^2 I) \\ Pr(\boldsymbol{\Psi}) &= Pr(\boldsymbol{\beta})Pr(\sigma^2)Pr(\tau^2)Pr(\phi). \end{aligned} \quad (5.9)$$

The common choices for the priors are the multivariate Normal for $\boldsymbol{\beta}$ and the inverse Gamma for σ^2 and τ^2 . The prior for ϕ depends on the choice of ρ function. In the simple exponential case, for example, where $\rho(\mathbf{s}_i - \mathbf{s}_j; \phi) = \exp(-\phi \|\mathbf{s}_i - \mathbf{s}_j\|)$, (with univariate ϕ), a Gamma prior is often chosen.

The *conditional* model (5.9) can be processed *hierarchically* as well, by writing the first-stage specification, regarding data conditional on both parameter and spatial process $f(\mathbf{Y} \mid \boldsymbol{\Psi}, \mathbf{W})$, and the second-stage specification, regarding the spatial process $Pr(\mathbf{W} \mid \boldsymbol{\theta})$ itself. The posterior distribution, $Pr(\boldsymbol{\Psi} \mid \mathbf{Y})$, in this case is proportional to $f(\mathbf{Y} \mid \boldsymbol{\Psi}, \mathbf{W})Pr(\mathbf{W} \mid \boldsymbol{\Psi})Pr(\boldsymbol{\Psi})$, i.e.,

$$Pr(\boldsymbol{\Psi} \mid \mathbf{Y}) \propto f(\mathbf{Y} \mid \boldsymbol{\Psi}, \mathbf{W})Pr(\mathbf{W} \mid \boldsymbol{\Psi})Pr(\boldsymbol{\Psi})$$

Formally, we can rewrite model (5.9) as

$$\begin{array}{ll} \text{first-stage} & \mathbf{Y} \mid \boldsymbol{\Psi}, \mathbf{W} \sim N_n(X\boldsymbol{\beta} + \mathbf{W}, \tau^2 I) \\ \text{second-stage} & \mathbf{W} \mid \sigma^2, \phi \sim N_n(\mathbf{0}, \sigma^2 H(\phi)), \end{array} \quad (5.10)$$

where $\mathbf{W} = (w(\mathbf{s}_1, \dots, \mathbf{s}_n))^T$ and such that $Y(\mathbf{s}_i)$ are conditionally independent given $w(\mathbf{s}_i)$, while $H(\phi)$ is the correlation matrix specified in (5.8). Even in this case, the Bayesian model specification is completed by adding priors for $\boldsymbol{\beta}$, τ^2 and for the hyperparameters σ^2 and ϕ . The conditional model is preferred because of its close form that makes the MCMC methods (the Gibbs sampling in particular) more efficient. Moreover, the first model is more steady. Note, in fact, that if two locations \mathbf{s}_i and \mathbf{s}_j are very close to each other, the matrix $\sigma^2 H(\phi)$ is close to singular while $\sigma^2 H(\phi) + \tau^2 I$ not.

5.4 Wrapped Spatial Models

Recalling the wrapped process definition in Chapter 3 and its density (3.5), we derive the wrapped spatial model as follows. Let $\mathbf{Y} = (Y_{s_1}, \dots, Y_{s_d})^T$ be a spatial Gaussian process defined on d spatial locations, i.e. $\mathbf{Y} \sim N_d(\tilde{\boldsymbol{\mu}}, \Sigma)$ where the spatial structure is specified through the covariance matrix Σ . To obtain the corresponding wrapped process, we apply to the vector \mathbf{Y} the wrapping procedure $X = Y(\text{mod} 2\pi)$ componentwise so that we achieve the circular vector $\mathbf{X} = (X_{s_1}, \dots, X_{s_d})^T$, where $Y_{s_i} = X_{s_i} + 2K_{s_i}\pi$ with $s \in \mathbf{s}, i = 1, \dots, d$, and each of the X_{s_i} is wrapped Normal $WN(\boldsymbol{\mu} = \tilde{\boldsymbol{\mu}}(\text{mod} 2\pi), \rho)$. Then, the spatial wrapped model can be completely specified given an appropriate definition of the spatial matrix Σ as shown in the previous subsection.

Considering as above a spatial model with nugget effect, then the covariance matrix is $\Sigma = \sigma^2 H(\phi) + \tau^2 I$ and the wrapped Normal spatial models is defined as follows:

$$\begin{aligned} \mathbf{X}, \mathbf{K} \mid \Psi &\sim N_d(\boldsymbol{\mu}, \sigma^2 H(\phi) + \tau^2 I) \\ Pr(\Psi) &= Pr(\boldsymbol{\mu}) Pr(\sigma^2) Pr(\tau^2) Pr(\phi). \end{aligned} \quad (5.11)$$

Note that model (5.11) is equivalent to the conditional model (5.9) assuming the equivalence $Y \equiv (X, K)$ that, as said in Section 3.1, has to be interpreted as the joint distribution for (X, K) with argument $(X + 2K\pi)$ and parameters vector $\Psi = \{\boldsymbol{\mu}, \sigma^2, \tau^2, \phi\}$. Moreover, expression (5.11) can be directly used into equations (3.9) and (3.14) in order to obtain the posterior distributions of all parameters and k 's coefficients of the wrapped spatial model.

That is, formally, we have

$$Pr(\Psi \mid \mathbf{X}, \mathbf{K}) \propto Pr(\mathbf{X}, \mathbf{K} \mid \Psi) Pr(\Psi), \quad (5.12)$$

and

$$Pr(\mathbf{K} \mid \mathbf{X}, \Psi) \propto Pr(\mathbf{X}, \Psi \mid \mathbf{K}) Pr(\mathbf{K}). \quad (5.13)$$

By equation (5.12) and (5.13) we have a fully specified Bayesian wrapped spatial model through which, in the next subsection, we compute spatial data examples. The first two applications regard simulated data while the last is a spatial data example applied to real data.

5.4.1 Examples

Simulated examples

Here we present two examples of simulated wrapped spatial processes with different dimension d . Simulation is straightforward. The in line spatial process, \mathbf{Y} , is simulated from a multivariate Normal distribution of dimension d and its corresponding wrapped process \mathbf{X} is obtained by applying the wrapping procedure componentwise: $\mathbf{X}^{(i)} = \mathbf{Y}^{(i)}(\text{mod } 2\pi)$, $i = 1 \dots, n$. This is repeated until a sample of size n is obtained. In our case, $n = 100$.

The likelihood used in (5.12) is:

$$P(\mathbf{X}, \mathbf{K} \mid \Psi = (\boldsymbol{\mu}, \Sigma)) \propto |\Sigma|^{-n/2} \exp \left\{ -\frac{1}{2} (\mathbf{X} + 2K\pi - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{X} + 2K\pi - \boldsymbol{\mu}) \right\} \quad (5.14)$$

where $\Sigma = (\sigma^2 H(\phi))$ is the covariance spatial matrix and H is the distance matrix. The parameters σ^2 and ϕ are the *partial sill* and the *range*, respectively. In these example we solve a wrapped Normal spatial model without nugget effect, modeled as a conditional model according to (5.9). Equation (5.14) is the likelihood of a "standard" in line Gaussian spatial process where the variable is $(X + 2K\pi)$.

The estimation procedure is the same described in Section 3.1.

For the first example, regarding a spatial process with dimension $d = 3$, we simulate a multivariate Normal distribution with vector mean $\boldsymbol{\mu} = \{0.7, 0.9, 1.2\}$ and spatial covariance matrix having partial sill $\sigma^2 = 0.8$ and range equal to 1.2.

Spatial wrapped Normal model (d=3)			
Parameter	True Value	Posterior Median	95% Posterior Interval
μ_1	0.80	0.76	[0.67, 0.98]
μ_2	1.00	1.04	[0.98, 1.21]
μ_3	1.20	1.28	[1.01, 1.38]
σ^2	0.80	0.82	[0.67, 0.95]
ϕ	1.2	1.26	[1.15, 1.43]

Table 5.3: Parameter estimates of the wrapped spatial Normal processes with dimension $d = 3$.

Spatial wrapped Normal model (d=6)			
Parameter	True Value	Posterior Median	95% Posterior Interval
μ_1	0.80	0.73	[0.63, 1.02]
μ_2	1.00	1.11	[0.89, 1.32]
μ_3	1.20	1.32	[1.06, 1.56]
μ_4	0.70	0.67	[0.61, 0.97]
μ_5	1.30	1.41	[1.16, 1.61]
μ_6	0.60	0.58	[0.53, 0.78]
σ^2	1.00	1.03	[0.91, 1.20]
ϕ	1.00	0.97	[0.86, 1.18]

Table 5.4: Parameter estimates of the wrapped spatial Normal processes with dimension $d = 6$.

Whereas for the second example with $d = 6$, we set the vector mean equal to $\boldsymbol{\mu} = \{0.8, 1, 1.2, 0.7, 1.3, 0.6\}$ and $\sigma^2 = \phi = 1$. For both simulated examples we set data on a regular 50km x50km lattice.

The priors used in the Metropolis-Hastings algorithm are: $\boldsymbol{\mu} \sim N_d(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$; $\sigma^2 \sim InvGamma(\alpha_0, \beta_0)$ and $\phi \sim Gamma(a_0, b_0)$. In particular we used the same setting in both the examples with $\boldsymbol{\mu}_0$ unit vector of dimension $d = 3$ and $d = 6$ respectively; and with $\alpha_0 = 4$, $\beta_0 = 8$, $a_0 = 3$ and $b_0 = 3$. In Figures 5.3-5.4 we show the traces (without burn in and thin) of the estimated parameter chains of Spatial WN Process with dimensions $d=3$ and $d=6$; while the complete inference results are reported in Tables 5.3 and 5.4.

Comparing Figures 5.3 and 5.4, the remarkable feature is the very different number of iterations need to reach convergence. When $d = 3$ the chains converge after about 1000 MCMC iterations. Instead, when the process dimension is $d = 6$, the simulated chains need more than 15000 iterations to converge. This evidence induces us to look for a general “behavior” of the Metropolis-Hastings (M-H) algorithm used for the wrapped spatial process. For this the reason we projected a simulation plan that takes into account several aspects of data and of the M-H algorithm itself as:

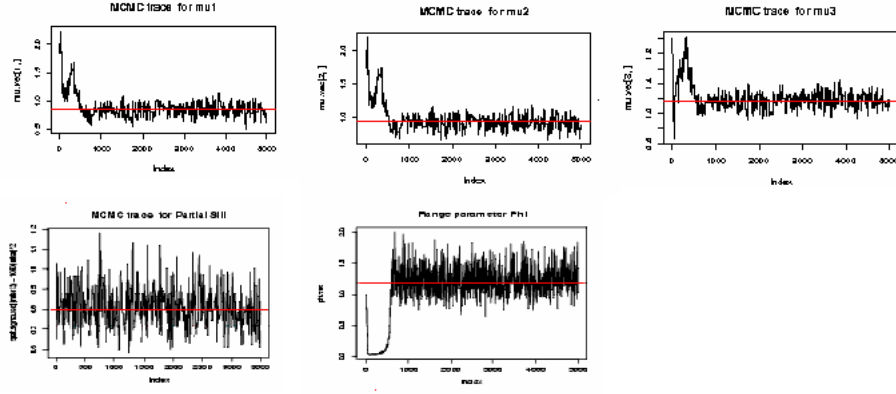


Figure 5.3: Traces of estimated parameters of Spatial WN Process with dimensions $d=3$

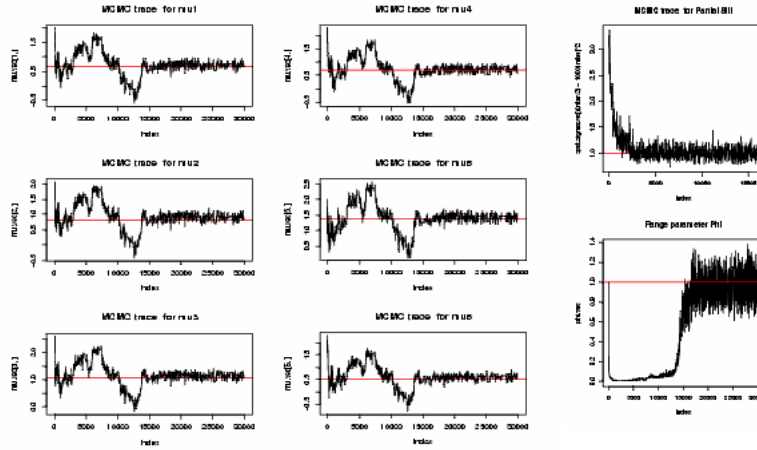


Figure 5.4: Traces of estimated parameters of Spatial WN Process with dimensions $d=6$

dimension of the process ($d=3,6,15,30$, so that 4 different simulation schemes are computed), *parameter setting* for transition density (a,b,c), *prior* setting (for μ , ϕ and σ^2), *variability* of the data (given by the variance of the true μ values) and *number of iterations*. As an example, we report in Table 5.5 the scheme concerning the simulation scheme for the spatial process with dimension $d = 3$. In appendix there are the schemes of the simulation plans regarding the spatial process with dimension $d = 6$, $d = 15$ and the results for each of them. Each of the 3 simulation schemes are constituted by 24 different simulations (the simulation identification is reported in the first column of the scheme), subdivided in 3 groups of 8 simulations. The simulated data are reported in column named “ μ true values ”. We choose two different setting for the data: data set with small variability $\mu = (-0.5, -0.2, 0.5)$ and data set with quite large variance $\mu = (-2, 0.4, 2)$. For both simulated data sets we choose different prior setting. In particular, for μ we differ in uncentred (about the mean of data) multivariate Normal distribution $MVN(0, \text{sigma.mu})$, and centred multivariate Normal distribution $MVN(-0.5, -0.2, 0.5)$. The kurtosis of these distribution is settled through the “Variance of μ Prior” setting, and also in this case we opt for two choices, more or less diffuse covariance matrix, obtained through “sigma.mu out-diag”, and “sigma.mu diag”.

Also for range parameter ϕ we set two different priors: $Gamma(3, 40)$ and $Gamma(2, 20)$, where the former is more informative than the latter. Follows the same aim we set two priors for partial sill parameter as well: $InvGamma(8, 9)$ and $InvGamma(4.8, 5)$. Finally, the setting for the transition density parameter are provided. Considering the Metropolis-Hastings algorithm illustrated in Section 3.1, we adopt a random walk as transition density, so a, b, c are the parameters of the uniform distributions related to the random walk errors for μ , σ^2 and ϕ respectively. Each of the 24 different simulations of this scheme is repeated 50 times and for a fix number of iterations equal to 5000.

Summarizing the results obtained by the simulation plans, we are able to detect some important characteristic and general behavior of spatial process and the M-H algorithm. First of all, when the process dimension increases, also the iteration numbers to reach chains convergence increases but non in a linear way. Moreover, analyzing the estimate accuracy (see estimates and posterior interval sizes related to simulation plans in appendix A) we confirm, (as just highlighted in Section 3.4) that large variability of the data leads to poor inference due to the identifiability problem. Finally, we discover that, as the computation complexity increases, the wrapped model becomes instable because of the large number of parameters to estimate. Note that for each variable of the process a k coefficient have to be estimated. So, another aspect, apart from the data variability, that involve inferential problem is the high dimension of the process. Notwithstanding these drawbacks, the posterior distribution for k ’s coefficients worked out in equation (3.14) allow us to overcome these problems.

As an example of the robustness of the inferential procedure described in Chapter 3, now we compute a high dimensional wrapped spatial model for real data.



Figure 5.5: Map of the 27 spatial location in the Adriatic sea

Real data example

For this example we use *Wam* data, provided by European Center for Medium-Range Weather and briefly described in first chapter. These data regard *wave mean direction* values estimated on 25km x 25km regular lattice on all the Mediterranean sea and recorded each 6 hour time interval. The data are located in a portion of Adriatic sea between Ancona and Ortona registered on December 19, 2002 at 00:00 am. The number of spatial locations considered is $d = 27$.

Figure 5.5 shows the geographic locations of the points in Adriatic sea whereas the wave mean direction data are reported in Table 5.6.

As in the previous simulated examples, we use the same prior setting for $\boldsymbol{\mu} \sim N_d(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$; $\sigma^2 \sim InvGamma(\alpha_0, \beta_0)$ and $\phi \sim Gamma(a_0, b_0)$, but, in this case, to set the prior parameters we use the information about data obtained whether from previous works or from simulation plan results. The Bayesian approach in this context of real data application is particularly suitable in that allows us to use the information about data directly in the inference procedure through opportune prior setting. In particular, from previous studies about these marine data (Bruschi *et al.*, 2005) we know that a reasonable value for μ_0 is around 0.5 and suitable priors for partial sill and range are $\sigma^2 \sim InvGamma(8, 9)$ and $\phi \sim Gamma(3, 40)$ respectively.

Another important inferential aspect regards the number of iterations of the M-H algorithm. For this high dimension model, we need at least 150000 iterations to reach the convergence and, in order to have uncorrelated chains, we choose a thin of 10. We report in Figure 5.6 the observed values, the estimated values and their 95% posterior interval for the vector parameter $\boldsymbol{\mu}$. The complete inferential results for this parameter vector are listed in Table 5.6.

Analyzing Figure 5.6, we note that the observed data are very close to estimate

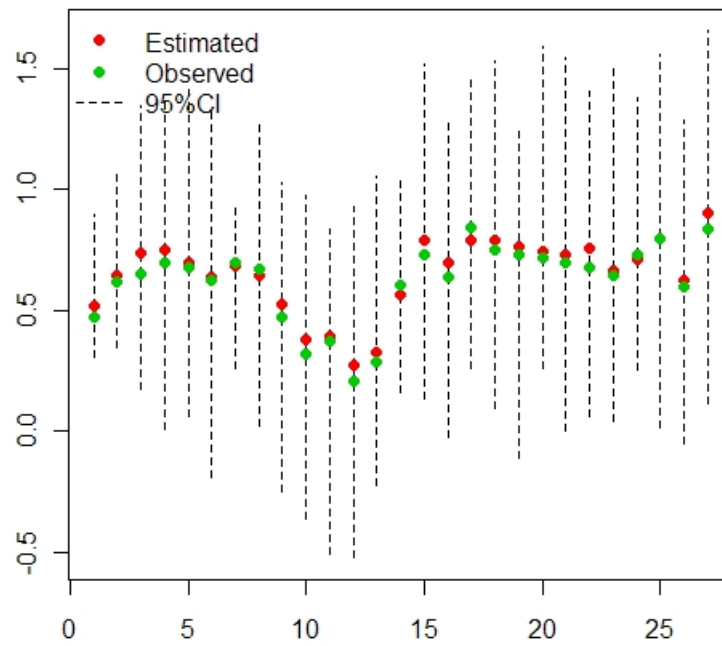


Figure 5.6: Observed values, estimated values and 95% posterior interval for the parameter vector μ

values even if the posterior interval sizes (quite large for all parameters) reveal not so accurate estimates. This is due to the high dimension that makes the model quite instable. In fact, comparing the posterior intervals sizes of the simulated example with $d = 6$ with posterior intervals sizes of model with $d = 3$, we note that in mean the second interval sizes are about a quarter of the first ones.

Probably, using more advanced MCMC algorithm we could obtain more quick and accurate results, but the main purpose here is providing a new method to apply the spatial models to circular data without taking into account of the computational aspects.

5.5 Discussion

In this chapter we introduce the spatial modeling. We provide general definitions of the spatial data categories and describe the common tools and notions of spatial statistics as stationarity, isotropy and variograms. Then, we focus on point-referenced data developing a Bayesian hierarchical model.

Next, extending the wrapped multivariate Normal density, we derive the wrapped spatial process formulation.

The ease and flexibility of the wrapping approach inference applied to spatial models is well represented in the examples that conclude the chapter. We first show two simulated examples with different number of spatial locations. The evidence that comes from these examples lead us to project a simulation plan in order to understand the behavior of the wrapping inferential procedure in a spatial context characterized by high dimensional models. The plan takes into account several aspects of both spatial model and Metropolis-Hastings algorithm. The knowledge obtained from simulation plan is directly used in inference procedure, through prior specification, to carry out the estimates of wrapped spatial model applied to real data set.

The reliability of the real data estimates constitute a good starting point to extend the wrapping inferential procedure to spatiotemporal models.

Simul identif	Transition density parameters		Prior on μ		Prior on ϕ (true=0.05)	Prior on σ^2 (true=1)	μ true values		Variance of μ prior	
	a	b	c				sigma.mu	out-diag	sigma.mu	diag
d3-1	1	1.3	2	MVN(0,sigma.mu)	gamma(3,40)	invgamma(8,9)	(-0.5,-0.2,0.5)	runif(0.5,1)	runif(0.5,1.6)	
d3-2	1	1.3	2	MVN(0,sigma.mu)	gamma(3,40)	invgamma(8,9)	(-0.5,-0.2,0.5)	runif(0.8,2)	runif(0.8,2.4)	
d3-3	1.8	1.3	2	MVN(0,sigma.mu)	gamma(3,40)	invgamma(8,9)	(-2,0.4,2)	runif(1,2.2)	runif(1,3.2)	
d3-4	1.8	1.3	2	MVN(0,sigma.mu)	gamma(3,40)	invgamma(8,9)	(-2,0.4,2)	runif(1,2.5)	runif(1,3.5)	
d3-5	1	1.3	2	MVN((-5,-2,-5),sigma.mu)	gamma(3,40)	invgamma(8,9)	(-0.5,-0.2,0.5)	runif(0.5,1)	runif(0.5,1.6)	
d3-6	1	1.3	2	MVN((-5,-2,-5),sigma.mu)	gamma(3,40)	invgamma(8,9)	(-0.5,-0.2,0.5)	runif(0.8,2)	runif(0.8,2.4)	
d3-7	1.8	1.3	2	MVN((-2,-4,2),sigma.mu)	gamma(3,40)	invgamma(8,9)	(-2,0.4,2)	runif(1,2.2)	runif(1,3.2)	
d3-8	1.8	1.3	2	MVN((-2,-4,2),sigma.mu)	gamma(3,40)	invgamma(8,9)	(-2,0.4,2)	runif(1,2.5)	runif(1,3.5)	
d3-9	1	1.3	2	MVN(0,sigma.mu)	gamma(2,20)	invgamma(8,9)	(-0.5,-0.2,0.5)	runif(0.5,1)	runif(0.5,1.6)	
d3-10	1	1.3	2	MVN(0,sigma.mu)	gamma(2,20)	invgamma(8,9)	(-0.5,-0.2,0.5)	runif(0.8,2)	runif(0.8,2.4)	
d3-11	1.8	1.3	2	MVN(0,sigma.mu)	gamma(2,20)	invgamma(8,9)	(-2,0.4,2)	runif(1,2.2)	runif(1,3.2)	
d3-12	1.8	1.3	2	MVN(0,sigma.mu)	gamma(2,20)	invgamma(8,9)	(-2,0.4,2)	runif(1,2.5)	runif(1,3.5)	
d3-13	1	1.3	2	MVN((-5,-2,-5),sigma.mu)	gamma(2,20)	invgamma(8,9)	(-0.5,-0.2,0.5)	runif(0.5,1)	runif(0.5,1.6)	
d3-14	1	1.3	2	MVN((-5,-2,-5),sigma.mu)	gamma(2,20)	invgamma(8,9)	(-0.5,-0.2,0.5)	runif(0.8,2)	runif(0.8,2.4)	
d3-15	1.8	1.3	2	MVN((-2,-4,2),sigma.mu)	gamma(2,20)	invgamma(8,9)	(-2,0.4,2)	runif(1,2.2)	runif(1,3.2)	
d3-16	1.8	1.3	2	MVN((-2,-4,2),sigma.mu)	gamma(2,20)	invgamma(8,9)	(-2,0.4,2)	runif(1,2.5)	runif(1,3.5)	
d3-17	1	1.3	2	MVN(0,sigma.mu)	gamma(3,40)	invgamma(4,8,5)	(-0.5,-0.2,0.5)	runif(0.5,1)	runif(0.5,1.6)	
d3-18	1	1.3	2	MVN(0,sigma.mu)	gamma(3,40)	invgamma(4,8,5)	(-0.5,-0.2,0.5)	runif(0.8,2)	runif(0.8,2.4)	
d3-19	1.8	1.3	2	MVN(0,sigma.mu)	gamma(3,40)	invgamma(4,8,5)	(-2,0.4,2)	runif(1,2.2)	runif(1,3.2)	
d3-20	1.8	1.3	2	MVN(0,sigma.mu)	gamma(3,40)	invgamma(4,8,5)	(-2,0.4,2)	runif(1,2.5)	runif(1,3.5)	
d3-21	1	1.3	2	MVN((-5,-2,-5),sigma.mu)	gamma(3,40)	invgamma(4,8,5)	(-0.5,-0.2,0.5)	runif(0.5,1)	runif(0.5,1.6)	
d3-22	1	1.3	2	MVN((-5,-2,-5),sigma.mu)	gamma(3,40)	invgamma(4,8,5)	(-0.5,-0.2,0.5)	runif(0.8,2)	runif(0.8,2.4)	
d3-23	1.8	1.3	2	MVN((-2,-4,2),sigma.mu)	gamma(3,40)	invgamma(4,8,5)	(-2,0.4,2)	runif(1,2.2)	runif(1,3.2)	
d3-24	1.8	1.3	2	MVN((-2,-4,2),sigma.mu)	gamma(3,40)	invgamma(4,8,5)	(-2,0.4,2)	runif(1,2.5)	runif(1,3.5)	

Table 5.5: Simulation plan scheme for wrapped spatial process with dimension $d = 3$

Id. Point	Coordinates X (UTM)	Coordinates Y(UTM)	Wave Direction	Estimate Wave direction	95% Posterior Interval
o1.1	458916.9	4705414	0.47	0.52	[0.31, 0.89]
o1.2	418491.4	4761300	0.62	0.64	[0.34, 1.08]
o1.3	438868.6	4761088	0.65	0.74	[0.17, 1.34]
o1.4	459245.8	4760936	0.70	0.75	[0.01, 1.37]
o1.5	479622.9	4760845	0.68	0.70	[0.06, 1.42]
o1.6	500000	4760815	0.62	0.64	[-0.18, 1.32]
o2.1	438621.3	4733326	0.70	0.69	[0.26, 0.94]
o2.2	459080.9	4733175	0.67	0.65	[0.02, 1.28]
o2.3	479540.5	4733084	0.47	0.53	[-0.25, 1.02]
o2.4	500000	4733053	0.32	0.38	[-0.35, 0.97]
o3.1	479458.5	4705323	0.38	0.39	[-0.51, 0.83]
o3.2	500000	4705293	0.21	0.28	[-0.52, 0.92]
o3.3	500000	4677534	0.29	0.33	[-0.22, 1.05]
o0.0	462149.4	4695402	0.61	0.56	[0.16, 1.04]
a1.1	360285.2	4901130	0.73	0.79	[0.14, 1.52]
a1.2	359694.5	4873362	0.64	0.70	[-0.02, 1.30]
a2.1	380244.6	4900734	0.84	0.79	[0.26, 1.45]
a2.2	379738.3	4872967	0.75	0.79	[0.09, 1.53]
a2.3	379234.4	4845200	0.73	0.76	[-0.11, 1.26]
a3.1	400203.9	4900400	0.72	0.74	[0.26, 1.59]
a3.2	399782.1	4872632	0.69	0.73	[0.01, 1.54]
a3.3	399362.1	4844866	0.68	0.76	[0.07, 1.40]
a4.1	419825.7	4872359	0.64	0.67	[0.04, 1.51]
a4.2	419489.8	4844592	0.73	0.71	[0.25, 1.37]
a4.3	419155.5	4816827	0.80	0.80	[0.02, 1.55]
a5.1	439617.4	4844380	0.60	0.63	[-0.05, 1.28]
a0.0	396280.2	4853800	0.83	0.91	[0.11, 1.65]

Table 5.6: Observed values, posterior estimates and posterior interval of the wave direction (expressed in radiant) on 25km x 25km lattice (UTM coordinates) recorded in a portion of Adriatic sea between Ancona and Ortona. The points identified by "o" are related to Ortona coast, while the points with "a" are related to Ancona coast.

Chapter 6

Spatiotemporal Modeling

In recent years there has been a tremendous growth in the statistical models and techniques to analyze spatiotemporal data such as air pollution data. Besides, the major part of the environmental data, as for example the rain precipitations, the wind directions, the temperatures and so on, are often observed both in spatial and in temporal contexts. It is not uncommon that the most important works about spatiotemporal modeling regard environmental data, see for example Wike and Cressie (1998), Wike (2003) and Gelfand *et al.* (2005).

Among the many possible approaches to the building of space-time models, the most natural in environmental science is the hierarchical one. Working with processes over bounded spatial and temporal domains, theoretical arguments tell us that we have few ergodicity results and, more generally, few situations where consistent estimation is possible. Hence, assessment of uncertainty through customary likelihood asymptotics can be a risky practice while a fully Bayesian analysis will result in exact inference. Admittedly, prior specification for spatio-temporal Bayesian models requires care and fitting will usually be more demanding (though the tools are becoming more and more available) but the resulting inferential comfort can provide justification.

In the next chapter we illustrate the basic framework of spatiotemporal modeling in the Bayesian hierarchical approach with particular attention to the dynamic formulation.

6.1 Spatiotemporal model formulation

The spatiotemporal modeling can be considered as an extension of spatial modeling of previous chapter. So, extending the definition of the spatial point-referenced model (5.7), a general form for point-referenced spatiotemporal model is

$$Y(\mathbf{s}, t) = \mu(\mathbf{s}, t) + e(\mathbf{s}, t) \quad \mathbf{s} = \mathbf{s}_1, \dots, \mathbf{s}_d \quad t = 1, \dots, T \quad (6.1)$$

where $Y(\mathbf{s}, t)$ denote the measurement at location \mathbf{s} and at time t ; $\mu(\mathbf{s}, t)$ denote the mean structure and $e(\mathbf{s}, t)$ are the residuals. As for the spatial model, even in this case, suppose $\mathbf{x}(\mathbf{s}, t)$ a vector of covariates associated to $Y(\mathbf{s}, t)$, we can express

the mean structure as $\mu(\mathbf{s}, t) = \mathbf{x}(\mathbf{s}, t)^T \beta(\mathbf{s}, t)$. Note that this last formulation allow spatiotemporal varying coefficient β 's even if usually it is adopted $\beta(\mathbf{s}, t) = \beta$, or only temporally or spatially varying coefficient as $\beta(\mathbf{s}, t) = \beta(t)$ and $\beta(\mathbf{s}, t) = \beta(\mathbf{s})$. Finally, $e(\mathbf{s}, t)$ can be written as $w(\mathbf{s}, t) + \xi(\mathbf{s}, t)$ where $\xi(\mathbf{s}, t)$ is a Gaussian white noise process and $w(\mathbf{s}, t)$ is the mean zero spatiotemporal process.

Again, as in spatial context, the model (6.1) can be viewed as conditional model given $\mu(\mathbf{s}, t)$ and $w(\mathbf{s}, t)$ at first stage, or hierarchically, given $w(\mathbf{s}, t)$ at second stage as well.

Therefore, all the spatiotemporal structure is given by $e(\mathbf{s}, t)$. Following Gelfand *et al.* (2004), the residuals are partitioned as follows:

$$e(\mathbf{s}, t) = \alpha(t) + w(\mathbf{s}) + \xi(\mathbf{s}, t), \quad (6.2)$$

$$e(\mathbf{s}, t) = \alpha_{\mathbf{s}}(t) + \xi(\mathbf{s}, t), \quad (6.3)$$

$$e(\mathbf{s}, t) = w_t(\mathbf{s}) + \xi(\mathbf{s}, t) \quad (6.4)$$

The expressions for $e(\mathbf{s}, t)$ do not invoke spatiotemporal interaction. Moreover, in each of (6.2), (6.3) and (6.4) the pure error terms are i.i.d $N(0, \sigma_{\xi}^2)$ and they represent a residual term of spatiotemporal explanation.

Now, we analyzing in detail the components of the previous three expressions (6.2)-(6.4), denoting with α all the temporal effects, while with w the spatial effects.

The first, (6.2), provide an additive temporal and spatial effect; while in (6.3), $\alpha_{\mathbf{s}}(t)$ indicates a temporal evolution at each site, and in (6.4) $w_t(\mathbf{s})$ indicates a spatial evolution over time. To better understand the structure of $\alpha_{\mathbf{s}}(t)$ and $w_t(\mathbf{s})$ we can collect the residuals $e(\mathbf{s}, t)$ into a $d \times T$ matrix where rows represent the spatial process over time and the columns represent the temporal process over space, i.e.,

$$e(\mathbf{s}, t) = \begin{pmatrix} e(s_1, 1), \dots, e(s_d, 1) \\ e(s_1, 2), \dots, e(s_d, 2) \\ \vdots \\ e(s_1, T), \dots, e(s_d, T) \end{pmatrix} = \begin{pmatrix} w_1(\mathbf{s}) \\ w_2(\mathbf{s}) \\ \vdots \\ w_T(\mathbf{s}) \end{pmatrix} = (\alpha_{s_1}(t), \alpha_{s_2}(t), \dots, \alpha_{s_d}(t))'. \quad (6.5)$$

For each time $t = 1, \dots, T$, $w_t(\mathbf{s})$ is a spatial process on locations s_1, \dots, s_d and, for each spatial location $s_i, i = 1, \dots, d$, $\alpha_{\mathbf{s}}(t)$ is a temporal process.

Returning to (6.2), $w(\mathbf{s})$ is modeled as a customary Gaussian spatial process following Section 5.3. For the temporal component, instead, if t is continuous, we can model $\alpha(t)$ as a stationary Gaussian process; in particular, posing, for a set (t_1, \dots, t_m) $\boldsymbol{\alpha} = (\alpha(t_1), \dots, \alpha(t_m))'$, we model $\boldsymbol{\alpha} \sim N(\mathbf{0}, \sigma_{\alpha}^2 \Sigma(\phi))$, where $(\Sigma(\phi))_{rs} = \rho(|t_r - t_s|; \phi)$ for a valid correlation function ρ . In case of discrete time points $t = 1, 2, \dots, T$, we can simply view $\alpha(1), \alpha(T)$ as the coefficients associated with a set of time dummy variables. With this assumption, we can adopt for $\alpha(t)$ a model as $\alpha(t) = \rho \alpha(t-1) + \epsilon(t)$ where $\epsilon(t)$ are i.i.d $N(0, \sigma_{\alpha}^2)$. In particular, if $\rho < 1$ we have a stationary autoregressive time series model $AR(1)$, while if $\rho = 1$ we model $\alpha(t)$ as random walk.

Considering the specification about $\alpha_s(t)$ mentioned above, the autoregressive or random walk model provide a suitable specification for the spatial component in (6.3). That is, we can assume $\alpha_s(t) = \rho\alpha_s(t-1) + \epsilon_s(t)$ where again $\epsilon_s(t)$ are i.i.d. The case with $\rho < 1$ is proposed in Hill *et al.* (1999) while the model for $\rho = 1$ can be found in Case and Shiller (1989).

Finally, the component $w_t(\mathbf{s})$ in (6.4) can be viewed as a collection of independent spatial process where the subscript t indicate the temporal slice at which the spatial process is referred. Note that with this expressions of $w_t(\mathbf{s})$ we are able to catch the temporal evolution of the spatial processes by comparing the spatial process parameters over time. Similar consideration about the spatial evolution of the temporal processes observed at each location are done.

6.2 Spatiotemporal Bayesian model specification

Now, we give the distribution specification for the model (6.1) for each of the cases (6.2), (6.3) and (6.4).

For a ease representation, we assume $\mathbf{Y}' = (\mathbf{Y}'_1, \dots, \mathbf{Y}'_T)$ where $\mathbf{Y}'_t = (Y(s_1, t), \dots, Y(s_d, t))$, and $\boldsymbol{\mu}' = (\mu'_1, \dots, \mu'_T)$, where $\mu_t = X_t\beta$ with X_t is the covariates matrix and β is supposed constant.

Under (6.2) let $\boldsymbol{\alpha}' = (\alpha(t_1), \dots, \alpha(t_T))$, $\mathbf{w}' = (\mathbf{w}(s_1), \dots, \mathbf{w}(s_n))$ and $\boldsymbol{\xi}' = (\xi(s_1, 1), \xi(s_1, 2), \dots, \xi(s_n, T))$. Then, the model (6.1) can be rewritten as

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\alpha} \otimes \mathbf{1}_{dx1} + \mathbf{1}_{Tx1} \otimes \mathbf{w} + \boldsymbol{\xi}, \quad (6.6)$$

where \otimes denote the Kronecker product. The corresponding conditional joint density distribution is

$$\mathbf{Y} \mid \beta, \boldsymbol{\alpha}, \mathbf{w}, \sigma_\xi^2 \sim N(\boldsymbol{\mu} + \boldsymbol{\alpha} \otimes \mathbf{1}_{dx1} + \mathbf{1}_{Tx1} \otimes \mathbf{w}, \sigma_\xi^2 I_{TdxTd}). \quad (6.7)$$

Conditionally to \mathbf{w} and $\boldsymbol{\alpha}$, \mathbf{Y}_t are independent, so the likelihood resulting from (6.7) arise as product of independent Normal densities. The posterior distribution can be obtained by the product of likelihood and opportune priors. The conditional model (6.7) is for sure the most intuitive to fit but it is characterized by a high-dimensional posterior distribution. Even in this case, just as for spatial model, we can obtain a much lower-dimensional posterior by marginalization. For this purpose, let $\mathbf{w} \sim N(\mathbf{0}, \sigma_w^2 H(\delta))$, and suppose that $\alpha(t)$ follows an $AR(1)$ so that $\boldsymbol{\alpha} \sim N(\mathbf{0}, \sigma_\alpha^2 A(\rho))$. Hence, if $\boldsymbol{\alpha}$, \mathbf{w} and $\boldsymbol{\xi}$ are independent, marginalizing over \mathbf{w} and $\boldsymbol{\alpha}$ we obtain the following marginal conditional density

$$\mathbf{Y} \mid \beta, \sigma_\xi^2, \sigma_\alpha^2, \rho, \sigma_w^2, \delta \sim N(\boldsymbol{\mu}, \sigma_\alpha^2 A(\rho) \otimes \mathbf{1}_{dx1} \mathbf{1}_{dx1}' + \sigma_w^2 \mathbf{1}_{Tx1} \mathbf{1}_{Tx1}' \otimes H(\delta) + \sigma_\xi^2 I_{TdxTd}). \quad (6.8)$$

Under (6.3), we have $\boldsymbol{\alpha}' = (\alpha(1)', \dots, \alpha(T)')$, where $\boldsymbol{\alpha}'(t) = (\alpha_{s_1}(t), \dots, \alpha_{s_d}(t))$. So the model (6.1) can be rewritten as

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\alpha} + \boldsymbol{\xi}.$$

The corresponding conditional joint density distribution is

$$\mathbf{Y} \mid \beta, \boldsymbol{\alpha}, \sigma_\xi^2 \sim N(\boldsymbol{\mu} + \boldsymbol{\alpha}, \sigma_\xi^2 I_{TdxTd}). \quad (6.9)$$

If the $\alpha_{s_i}(t)$ follow a $AR(1)$ model independently across i , then, marginalizing over $\boldsymbol{\alpha}$ we obtain the following marginal conditional density:

$$\mathbf{Y} \mid \beta, \sigma_\xi^2, \sigma_\alpha^2, \rho, \sim N(\boldsymbol{\mu}, A(\rho) \otimes I_{TdxTd} + \sigma_\xi^2 I_{TdxTd}). \quad (6.10)$$

In a similar way, we can rewrite the model (6.1) under the assumption (6.4), assuming $\mathbf{w}' = (\mathbf{w}'_1, \dots, \mathbf{w}'_T)$ where $\mathbf{w}'_t = (w_t(s_1), \dots, w_t(s_d))$, as

$$\mathbf{Y} = \boldsymbol{\mu} + \mathbf{w} + \boldsymbol{\xi},$$

and the corresponding conditional joint density distribution is

$$\mathbf{Y} \mid \beta, \mathbf{w}, \sigma_\xi^2 \sim N(\boldsymbol{\mu} + \mathbf{w}, \sigma_\xi^2 I_{TdxTd}). \quad (6.11)$$

If the $\mathbf{w} \sim N(\mathbf{0}, \sigma_w^{2(t)} H(\delta^{(t)}))$ independently across $t = 1, \dots, T$, then, marginalizing over \mathbf{w} we obtain the following marginal conditional density:

$$\mathbf{Y} \mid \beta, \sigma_\xi^2, \sigma_w^2, \boldsymbol{\delta}, \sim N(\boldsymbol{\mu}, D(\boldsymbol{\sigma}_w^2, \boldsymbol{\delta}) + \sigma_\xi^2 I_{TdxTd}), \quad (6.12)$$

where $\boldsymbol{\sigma}_w^{2'} = (\sigma_w^{2(1)}, \dots, \sigma_w^{2(T)})$, $\boldsymbol{\delta}' = (\delta^{(1)}, \dots, \delta^{(T)})$ and $D(\boldsymbol{\sigma}_w^2, \boldsymbol{\delta})$ is a block diagonal matrix with the t^{th} block given by $\sigma_w^{2(t)} H(\delta^{(t)})$. Because D is block diagonal, likelihood evaluation associated with (6.12) is easier than for (6.8).

Finally, note that with either (6.3) or (6.4), $e(\mathbf{s}, t)$ is comprised of two sources of error that the data cannot directly separate. However, through the assumption on the $\alpha_{\mathbf{s}}(t)$ or on the $w_t(\mathbf{s})$, we can learn about the process that guide the error components as (6.10) and (6.12) reveal.

More details and results can be found in Banerjee *et al.* (2004, Chap. 8).

6.3 Dynamic Spatiotemporal modeling

In this section we present an approach for data that arise from a time series of spatial processes. We decide to work in the setting of Bayesian dynamic models (West and Harrison, 1997). In the following we first give the basic assumption and definition for dynamic linear modeling, then we describe the temporal evolution of a latent space model by a dynamic spatiotemporal model. This notation and theoretical framework are necessary to implement the dynamical spatio-temporal model in next sections.

6.3.1 Dynamic linear model

Dynamic linear model often referred to as state-space models in time-series literature, offer a versatile framework for fitting several time-varying models. Here we

describe directly the multivariate dynamic linear model (DLM) considering the univariate DLM as a special case of the former.

Suppose for $t = 1, \dots, T$ that \mathbf{Y}_t is a (column) vector of d observations on the time series $\mathbf{Y}_1, \dots, \mathbf{Y}_T$. The multivariate DLM is defined via a quadruple

$$\{\mathbf{F}, \mathbf{G}, \mathbf{V}, \mathbf{W}\}_t = \{\mathbf{F}_t, \mathbf{G}_t, \mathbf{V}_t, \mathbf{W}_t\}$$

for each time t , where:

\mathbf{F}_t is a known (pxd) dynamic regression matrix;

\mathbf{G}_t is a known (pxp) state evolution matrix;

\mathbf{V}_t is a known (dxd) observational variance matrix;

\mathbf{W}_t is a known (pxp) evolution variance matrix.

The corresponding model equations are:

$$\mathbf{Y}_t = \mathbf{F}_t' \boldsymbol{\theta}_t + \boldsymbol{\nu}_t \quad \boldsymbol{\xi}_t \sim N(\mathbf{0}, \mathbf{V}_t) \quad \text{measurement equation,} \quad (6.13)$$

$$\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t \quad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \mathbf{W}_t) \quad \text{transition equation,} \quad (6.14)$$

where the error sequences $\boldsymbol{\nu}_t$ and $\boldsymbol{\omega}_t$ are independent and mutually independent. Equation (6.13) is the *measurement equation* for the model, defining the sampling distribution for \mathbf{Y}_t conditional on the quantity $\boldsymbol{\theta}_t$. It is assumed that, given this vector, \mathbf{Y}_t is conditionally independent of the past values of the series. This equation relates the observations to $\boldsymbol{\theta}_t$ via a dynamic linear regression with a multivariate normal error structure having known, though possibly time varying, observation variance matrix \mathbf{V}_t . The matrix \mathbf{F}_t plays the role of the *regression matrix* of known values of independent variables, and $\boldsymbol{\theta}_t$ is the dynamic vector of the regression parameters, referred to as the *state vector* of the model. The *mean response* at t is $\boldsymbol{\mu}_t = \mathbf{F}_t' \boldsymbol{\theta}_t$, simply the expected value of the \mathbf{Y}_t in (6.13), which defines the level of the series at time t . Finally, in the observation equation, $\boldsymbol{\xi}_t$ is the observational error at time t . Equation (6.14) is the *state transition equation*, defining the time evolution of the state vector. A one-step Markov evolution is evident; given $\boldsymbol{\theta}_{t-1}$, and the known values of the \mathbf{G}_t and \mathbf{W}_t , the distribution of $\boldsymbol{\theta}_t$ is fully determined independently of values of the state vector and data prior to time $t - 1$. The deterministic component of the evolution is the transition from state $\boldsymbol{\theta}_{t-1}$ to $\mathbf{G}_t \boldsymbol{\theta}_{t-1}$, a simply linear transformation via the state *transition transfer matrix* \mathbf{G}_t . The evolution is completed with the addition of the *transition error* term $\boldsymbol{\epsilon}_t$, with known transition variance matrix \mathbf{W}_t .

Note that the $\boldsymbol{\xi}$ error is simply a random permutation in the measurement process that affects the observations \mathbf{Y}_t but has no further influence on the series. The error $\boldsymbol{\epsilon}_t$, by contrast, influences the development of the system into the future.

6.3.2 Updating equation

Now we show the recursive procedure that allows the updating measurement and state equation of the model DLM (6.13)-(6.14).

The central characteristic of the DLM (6.13)-(6.14) is that, at any time, existing information about the system is represented and sufficiently summarized by the posterior distribution for the current state vector. These posterior distributions are generally classified as:

predictive distribution $\boldsymbol{\theta}_t \mid \mathbf{Y}_{t-1}, \dots, \mathbf{Y}_1$: the state at current time is conditioned to information up to previous times;

filter distribution $\boldsymbol{\theta}_t \mid \mathbf{Y}_1, \dots, \mathbf{Y}_t$: the state at current time is conditioned to information up to current times;

smoother distribution $\boldsymbol{\theta}_t \mid \mathbf{Y}_T, \mathbf{Y}_{T-1}, \dots, \mathbf{Y}_1$: the state at current time is conditioned to *all* information.

In particular the predictive distribution is

$$\boldsymbol{\theta}_t \mid \mathbf{Y}_{t-1}, \dots, \mathbf{Y}_1 \sim N_p(\mathbf{a}_t, \mathbf{R}_t)$$

where

$$\mathbf{a}_t = \mathbf{G}_t \mathbf{m}_{t-1} \quad \text{and} \quad \mathbf{R}_t = \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t' + \mathbf{W}_t; \quad (6.15)$$

the filtering distribution is

$$\boldsymbol{\theta}_t \mid \mathbf{Y}_1, \dots, \mathbf{Y}_t \sim N_p(\mathbf{m}_t, \mathbf{C}_t)$$

with

$$\mathbf{m}_t = \mathbf{a}_t + \mathbf{A}_t \mathbf{e}_t \quad \text{and} \quad \mathbf{C}_t = \mathbf{R}_t - \mathbf{A}_t \mathbf{Q}_t \mathbf{A}_t' \quad (6.16)$$

where

$$\mathbf{A}_t = \mathbf{R}_t \mathbf{F}_t \mathbf{Q}_t^{-1} \quad \text{and} \quad \mathbf{e}_t = \mathbf{Y}_t - \mathbf{f}_t$$

and

$$\mathbf{Q}_t = \mathbf{F}_t' \mathbf{R}_t \mathbf{F}_t + \mathbf{V}_t \quad \text{and} \quad \mathbf{f}_t = \mathbf{F}_t' \mathbf{a}_t.$$

Finally, the smoother distribution is given by

$$\boldsymbol{\theta}_T \mid \mathbf{Y}_T, \mathbf{Y}_{T-1}, \dots, \mathbf{Y}_1 \sim N_p(\mathbf{m}_T, \mathbf{C}_T).$$

The opportune combinations of these posterior distributions allow us to obtain the posterior estimate of the state vector $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_T)$ as follows.

Pose, $\mathbf{Y} = (\mathbf{Y}_1, \dots, \mathbf{Y}_T)$ from the Markov structure of the DLM it is demonstrated that the posterior distribution of the state vector is

$$Pr(\boldsymbol{\Theta} \mid \mathbf{Y}) = Pr(\boldsymbol{\theta}_T \mid \mathbf{Y}) \prod_{t=1}^{T-1} Pr(\boldsymbol{\theta}_t \mid \mathbf{Y}_1, \dots, \mathbf{Y}_t, \boldsymbol{\theta}_{t+1}).$$

So, in order to generate $\boldsymbol{\Theta}$ from $Pr(\boldsymbol{\Theta} \mid \mathbf{Y})$ we backward sample from smoother distribution as follows:

(a) for $t = T$ we sample $\boldsymbol{\theta}_T$

$$\boldsymbol{\theta}_T \mid \mathbf{Y} \sim N(\mathbf{m}_T, \mathbf{C}_T)$$

(b) for $t = T - 1, \dots, 1$ we sample $\boldsymbol{\theta}_t$ from $Pr(\boldsymbol{\theta}_t \mid \mathbf{Y}_1, \dots, \mathbf{Y}_t, \boldsymbol{\theta}_{t+1})$, where

$$\boldsymbol{\theta}_t \mid \mathbf{Y}_1, \dots, \mathbf{Y}_t, \boldsymbol{\theta}_{t+1} \sim N(\mathbf{h}_t, \mathbf{H}_t)$$

with

$$\mathbf{h}_t = \mathbf{m}_t + \mathbf{B}_t(\boldsymbol{\theta}_{t+1} - \mathbf{a}_{t+1}) \quad \text{and} \quad \mathbf{H}_t = \mathbf{C}_t - \mathbf{B}_t \mathbf{R}_{t+1} \mathbf{B}_t'$$

and

$$\mathbf{B}_t = \mathbf{C}_t \mathbf{G}_{t+1}' \mathbf{R}_{t+1}^{-1}.$$

The vectors $\mathbf{m}_t, \mathbf{a}_t, \mathbf{C}_t, \mathbf{R}_t$ ($t = 1, \dots, T$) are the output of the *forward filtering*. In particular, starting from $\boldsymbol{\theta}_0 \sim N(\mathbf{m}_0, \mathbf{C}_0)$ we compute $\boldsymbol{\theta}_t$ by the prediction distribution

$$\boldsymbol{\theta}_t \mid \mathbf{Y}_1, \dots, \mathbf{Y}_{t-1} \sim N(\mathbf{a}_t, \mathbf{R}_t),$$

with \mathbf{a}_t and \mathbf{R}_t , defined in (6.15); then we draw $\boldsymbol{\theta}_t$ from the filter distribution

$$\boldsymbol{\theta}_t \mid \mathbf{Y}_1, \dots, \mathbf{Y}_t \sim N(\mathbf{m}_t, \mathbf{C}_t),$$

where \mathbf{m}_t and \mathbf{C}_t are defined in (6.16).

Hence, in the forward sampling, starting from $\boldsymbol{\theta}_0 \sim N(\mathbf{m}_0, \mathbf{C}_0)$ (\mathbf{m}_0 and \mathbf{C}_0 fixed) for each $t = 1, \dots, T$ we sample from

$$\boldsymbol{\theta}_t \mid \mathbf{Y}_1, \dots, \mathbf{Y}_t \sim N(\mathbf{m}_t, \mathbf{C}_t),$$

and we save $\mathbf{a}_t, \mathbf{R}_t$ through equations (6.15). In the *backward sampling*, starting from $\boldsymbol{\theta}_T \mid \mathbf{Y} \sim N(\mathbf{m}_T, \mathbf{C}_T)$, for each $t = T - 1, \dots, 1$ we sample from

$$\boldsymbol{\theta}_t \mid \mathbf{Y}_1, \dots, \mathbf{Y}_t, \boldsymbol{\theta}_{t+1} \sim N(\mathbf{h}_t, \mathbf{H}_t).$$

6.3.3 Dynamic formulation for spatiotemporal models

In this subsection we adapt the above dynamic modeling framework to univariate spatiotemporal models (defined in 6.1) with spatially varying coefficients.

The response $\mathbf{Y}(\mathbf{s}, t)$ is first modeled through a measurement equation, which incorporate the measurement error $\xi(\mathbf{s}, t)$, as serially and spatially uncorrelated zero-mean Gaussian disturbances. The transition equation involves the regression parameters of the covariates. Here we adopt a very flexible formulation for the spatiotemporal model, originally provided in Gelfand *et al.* (2005), where the regression parameters $\tilde{\beta}(\mathbf{s}, t)$ is decomposed into a purely temporal component, β_t , and a spatiotemporal component $\beta(\mathbf{s}, t)$. Both of them are generated through transition equation, capturing their Markovian dependence in time. The transition equation of the purely

temporal component is as in usual DLM, while the spatiotemporal component is generated by a multivariate Gaussian spatial process. Thus, we can write the spatiotemporal modeling framework as

Measurement equation:

$$\begin{aligned} \mathbf{Y}(\mathbf{s}, t) &= \boldsymbol{\mu}(\mathbf{s}, t) + \boldsymbol{\xi}(\mathbf{s}, t) \quad \boldsymbol{\xi} \sim N_d(\mathbf{0}, \sigma_\xi^2), \\ \boldsymbol{\mu}(\mathbf{s}, t) &= \mathbf{x}^T(\mathbf{s}, t) \tilde{\boldsymbol{\beta}}(\mathbf{s}, t) \\ \tilde{\boldsymbol{\beta}}(\mathbf{s}, t) &= \beta_t + \boldsymbol{\beta}(\mathbf{s}, t), \end{aligned} \quad (6.17)$$

Transition equation:

$$\begin{aligned} \beta_t &= \beta_{t-1} + \epsilon_t, \quad \epsilon_t \sim N_p(\mathbf{0}, \Sigma_\epsilon), \\ \boldsymbol{\beta}(\mathbf{s}, t) &= \boldsymbol{\beta}(\mathbf{s}, t-1) + \mathbf{w}(\mathbf{s}, t). \end{aligned} \quad (6.18)$$

In (6.18) we introduce a linear model of coregionalization (Banerjee *et al.*, 2004, Sec. 7.2) for $\mathbf{w}(\mathbf{s}, t)$, i.e., $\mathbf{w}(\mathbf{s}, t) = \mathbf{A}\mathbf{v}(\mathbf{s}, t)$ with $\mathbf{v}(\mathbf{s}, t)$ vector of p independent replications of a Gaussian spatial process. Following subsection 6.3.1, the Bayesian hierarchical model for (6.17) and (6.18) may be completed by prior specification as

$$\begin{aligned} \beta_0 &\sim N(\mathbf{m}_0, \mathbf{C}_0) \\ \Sigma_\epsilon &\sim IW(a_\xi, B_\xi) \\ \Sigma_{\mathbf{w}} &\sim IW(a_{\mathbf{w}}, B_{\mathbf{w}}) \\ \sigma_\xi^2 &\sim IG(a_\xi, b_\xi) \\ \mathbf{m}_0 &\sim N(\mathbf{0}, \Sigma_0) \end{aligned} \quad (6.19)$$

In order to achieve a DLM representation of the model (6.17)-(6.18), we collect, for each time point, the observations on all sites into $\mathbf{Y}_t = (Y(s_1, t), \dots, Y(s_d, t))^T$. Consequently, also the covariates $\mathbf{F}_t = (\mathbf{x}^T(s_1, t), \dots, \mathbf{x}^T(s_d, t))$ and the temporal, β_t , and spatiotemporal, $\boldsymbol{\beta}(\mathbf{s}, t)$, regression parameters $\boldsymbol{\theta}_t = \mathbf{1}_d \otimes \beta_t + \boldsymbol{\beta}_t^*$ are collected, so we can write the data equation of a the dynamic spatial model as a usual DLM measurement equation form:

$$\begin{aligned} \mathbf{Y}_t &= \mathbf{F}_t \boldsymbol{\theta}_t + \boldsymbol{\xi}_t \quad \boldsymbol{\xi}_t \sim N_d(\mathbf{0}, \sigma_\xi^2 \mathbf{I}) \\ \boldsymbol{\theta}_t &= \mathbf{1}_d \otimes \beta_t + \boldsymbol{\beta}_t^*, \end{aligned} \quad (6.20)$$

where $\boldsymbol{\beta}_t^* = (\beta(s_1, t), \dots, \beta(s_d, t))^T$.

With the prior specification in (6.19) the model (6.20) is fully specified. For more detail see e.g. Gelfand *et al.* (2005).

6.4 Wrapped Spatiotemporal Modeling

The wrapped spatiotemporal model is an extension of the wrapped spatial model and it can be obtained in a similar way.

Given the general form of the spatiotemporal model $Y(\mathbf{s}, t) = \mu(\mathbf{s}, t) + e(\mathbf{s}, t)$ $\mathbf{s} =$

$\mathbf{s}_1, \dots, \mathbf{s}_d \quad t = 1, \dots, T$, we can apply the wrapping procedure for each additive component of the model. In fact, the expression (6.1), can be viewed as a measurement error model and then we can resort to the approach illustrated in Section 4.4 to build a spatiotemporal wrapped model.

Without lose in generalization, under the assumption (6.2), we can rewrite the model (6.1) as follows:

$$\begin{aligned} Y(\mathbf{s}, t) &= \mu(\mathbf{s}, t) + \alpha(t) + w(\mathbf{s}) + \xi(\mathbf{s}, t) \quad \mathbf{s} = (s_1, \dots, s_d) \quad t = 1, \dots, T \\ &= \chi(\mathbf{s}, t) + \xi(\mathbf{s}, t) \\ \text{with } \chi(\mathbf{s}, t) &= \mu(\mathbf{s}, t) + \alpha(t) + w(\mathbf{s}) \end{aligned} \tag{6.21}$$

The last two equation in (6.21) are clearly represented in a ME Berkson model form (Section 4.1). Thus, as illustrated in Section 4.4, we can again resort to the distribution property (a) of the wrapping distribution in Section 2.4.3, $(y_1 + y_2)_w = y_{1w} + y_{2w}$, in order to obtain the results as in (4.19).

A legitimate objection is that in this context, $\chi(\mathbf{s}, t)$ is not a variable as in Berkson ME model but a process. Then, again we can apply the wrapping procedure $X_i = Y_i(\text{mod} 2\pi)$ componentwise to each in line component Y_i of the temporal or spatial process as shown in Section 5.4.

Without loss of generalization, consider the expression (6.21) with a mean structure, $\mu(\mathbf{s}, t) = \mathbf{x}(\mathbf{s}, t)\beta$, constant and centered on zero, the wrapped spatiotemporal model is given by

$$\begin{aligned} (Y(\mathbf{s}, t))_w &= (\chi(\mathbf{s}, t) + \xi(\mathbf{s}, t))_w \\ &= \alpha_w(t) + w_w(\mathbf{s}) + \xi_w(\mathbf{s}, t) \end{aligned}$$

Moreover, consider a time continuous process and let $\alpha(t) = (\alpha(t_1), \dots, \alpha(t_T))'$ be a stationary Gaussian process. In this case we apply the wrapping procedure componentwise to $\alpha(t_i) \quad i = 1, \dots, T$ in order to obtain the corresponding wrapped Normal process distributed as $WN_T(\mathbf{0}, \sigma_\alpha^2 \Sigma(\phi))$. In the same way, supposing $w(\mathbf{s})$ be a d -dimensional spatial Gaussian process $N(\mathbf{0}, \sigma_w^2 \Sigma(\phi_w))$ and $\xi(s, t) \sim N(0, \sigma_\xi^2)$, we can achieve their corresponding wrapped distribution as follows:

$$\begin{aligned} \alpha_w(t) &\sim WN_T(\mathbf{0}, \sigma_\alpha^2 \Sigma(\phi_\alpha)) \\ w(\mathbf{s}) &\sim WN_d(\mathbf{0}, \sigma_w^2 \Sigma(\phi_w)) \\ \xi(\mathbf{s}, t) &\sim WN(0, \sigma_\xi^2) \end{aligned} \tag{6.22}$$

Through the expression (6.21) and with the specifications (6.22) the spatiotemporal model under the assumption (6.2) is fully derived. In a similar way we can easily derive the spatiotemporal model under the assumptions (6.3) and (6.4).

6.4.1 Simulated Example

Here, we aim to provide an hierarchically development of a ME model. Our spatiotemporal modeling framework is given by:

$$X_{obs}(\mathbf{s}, t) = X_{true}(\mathbf{s}, t) + \eta_{obs}(\mathbf{s}, t) \quad (6.23)$$

$$X_{true}(\mathbf{s}, t) = \alpha X_{true}(\mathbf{s}, t-1) + \eta_{true}(\mathbf{s}, t) \quad (6.24)$$

where $X_{obs}(\mathbf{s}, t)$, $\eta_{obs}(\mathbf{s}, t)$ and $\eta_{true}(\mathbf{s}, t)$ are supposed distributed as wrapped Normal and $X_{true}(\mathbf{s}, t)$ is a wrapped autoregressive process (see Breckling, 1989).

Now, following the wrapping approach inference (Section 3.1), the equivalence $Y \equiv (X, K)$ holds, $((X, K)$ is the joint distribution with argument $(X + 2K\pi)$ and then, posing $Y_{obs} = (X_{obs} + 2K_{obs}\pi)$ and $Y_{true} = (X_{true} + 2K_{true}\pi)$, the model (6.23)-(6.24) is rewritten as:

$$Y_{obs}(\mathbf{s}, t) = Y_{true}(\mathbf{s}, t) + \xi(\mathbf{s}, t) \quad \xi(\mathbf{s}, t) \sim N(0, \sigma_\xi^2) \quad (6.25)$$

$$Y_{true}(\mathbf{s}, t) = \alpha Y_{true}(\mathbf{s}, t-1) + \epsilon(\mathbf{s}, t) \quad \epsilon(\mathbf{s}, t) \sim N(0, \sigma_\epsilon^2) \quad (6.26)$$

Collecting the observations over spatial locations, we obtain $\mathbf{Y}_{obs}(t) = (Y_{obs}(\mathbf{s}_1, t), \dots, Y_{obs}(\mathbf{s}_d, t))$, $\mathbf{Y}_{true}(t) = (Y_{true}(\mathbf{s}_1, t), \dots, Y_{true}(\mathbf{s}_d, t))$, $\boldsymbol{\xi}_s(t) = (\xi(\mathbf{s}_1, t), \dots, \xi(\mathbf{s}_d, t))$ and $\boldsymbol{\epsilon}_s(t) = (\epsilon(\mathbf{s}_1, t), \dots, \epsilon(\mathbf{s}_d, t))$. With this formulation, the model (6.25)-(6.26) can be rewritten as follows:

$$\mathbf{Y}_{obs}(t) = \mathbf{Y}_{true}(t) + \boldsymbol{\xi}_s(t) \quad \boldsymbol{\xi}_s(t) \sim N_d(0, \sigma_\xi^2 I) \quad (6.27)$$

$$\mathbf{Y}_{true}(t) = \alpha \mathbf{Y}_{true}(t-1) + \boldsymbol{\epsilon}_s(t) \quad \boldsymbol{\epsilon}_s(t) \sim N_d(0, \Sigma_\epsilon(t)) \quad (6.28)$$

where $\Sigma_\epsilon(t)$ is the covariance matrix that characterize the spatial structure of data. The model (6.27)-(6.28) presents a usual DLM form where (6.27) is the *measurement equation* while (6.28) is the *transition equation* for the *(latent) dynamic spatial Gaussian process* $\mathbf{Y}_{true}(t)$. Thus, the latent process is computed by inferential procedure shown in Section 6.3.

The fully Bayesian specification involves the likelihood and the prior specifications in order to obtain the posterior distribution. For the model (6.27)-(6.28), assuming $\boldsymbol{\Psi} = (\sigma_\xi^2, \alpha, \Sigma_\epsilon(t))$, the posterior distribution is:

$$Pr(\boldsymbol{\Psi}, \mathbf{Y}_{true} \mid \mathbf{Y}_{obs}) \propto Pr(\mathbf{Y}_{obs} \mid \boldsymbol{\Psi}, \mathbf{Y}_{true}) Pr(\mathbf{Y}_{true} \mid \boldsymbol{\Psi}) Pr(\boldsymbol{\Psi}) \quad (6.29)$$

where

$$\mathbf{Y}_{obs} \mid \boldsymbol{\Psi}, \mathbf{Y}_{true} \equiv \mathbf{Y}_{obs}(t) \mid \mathbf{Y}_{true}(t), \sigma_\xi^2 \sim N_d(\mathbf{Y}_{true}(t), \sigma_\xi^2 I) \quad (6.30)$$

and

$$\mathbf{Y}_{true} \mid \boldsymbol{\Psi} \equiv \mathbf{Y}_{true}(t) \mid \alpha \mathbf{Y}_{true}(t-1), \Sigma_\epsilon(t) \sim N_d(\alpha \mathbf{Y}_{true}(t-1), \Sigma_\epsilon(t)) \quad (6.31)$$

Returning to the wrapped formulation, the corresponding wrapped model of (6.29) is:

$$\begin{aligned} Pr(\Psi, \mathbf{X}_{true}, \mathbf{K}_{true} \mid \mathbf{X}_{obs}, \mathbf{K}_{obs}) &\propto \\ Pr(\mathbf{X}_{obs}, \mathbf{K}_{obs} \mid \Psi, \mathbf{X}_{true}, \mathbf{K}_{true}) &Pr(\mathbf{X}_{true}, \mathbf{K}_{true} \mid \Psi) Pr(\Psi) \end{aligned} \quad (6.32)$$

where

$$\mathbf{X}_{obs}, \mathbf{K}_{obs} \mid \Psi, \mathbf{X}_{true}, \mathbf{K}_{true} \sim N_d(\mathbf{X}_{true}(t) + 2\mathbf{K}_{true}\pi, \sigma_\xi^2 I) \quad (6.33)$$

and

$$\mathbf{X}_{true}, \mathbf{K}_{true} \mid \Psi \sim N_d(\alpha(\mathbf{X}_{true}(t-1) + 2\mathbf{K}_{true}(t-1)\pi), \Sigma_\epsilon(t)) \quad (6.34)$$

The wrapped spatiotemporal model is completely specified adding the posterior distribution for k 's wrapping coefficients, as follows:

$$Pr(\mathbf{K}_* \mid \mathbf{X}_* \Psi) \propto Pr(\mathbf{X}_*, \Psi \mid \mathbf{K}_*) Pr(\mathbf{K}_*)$$

where $*$ means that the expression holds for both \mathbf{K}_{obs} and \mathbf{K}_{true} . Parameter estimates for the model (6.27)-(6.28) are obtained through the following two steps:

- A. A Metropolis step, as illustrated in Section 3.1, for each parameter of $\Psi = (\sigma_\xi^2, \alpha, \Sigma_\epsilon(t))$;
- B. A forward filtering backward sampling (FFBS) step, as described in Section 6.3, to compute $\mathbf{Y}_{true}(t)$, $t = 1, \dots, T$

Now we see the generation of the simulated data in more detail. Considering the model (6.27)-(6.28), as usual, we simulate in line data, $\mathbf{Y}_{obs}(t)$ and $\mathbf{Y}_{true}(t)$, and then apply the wrapping procedure $X = Y(mod 2\pi)$ to obtain circular data, $\mathbf{X}_{obs}(t)$ and $\mathbf{X}_{true}(t)$. Here, $\mathbf{Y}_{obs}(t)$ and $\mathbf{Y}_{true}(t)$ are spatial processes, thus to obtain their corresponding circular process we need to apply wrapped procedure for each of the d components, $(Y_{obs}(\mathbf{s}_1, t), \dots, Y_{obs}(\mathbf{s}_d, t))$, as illustrated in Section 5.4.

$\mathbf{Y}_{obs}(t)$ and $\mathbf{Y}_{true}(t)$ are both drawn from d dimensional Normal distribution. In particular, $\mathbf{Y}_{obs}(t) \sim N_d(\mathbf{Y}_{true}(t), \sigma_\xi^2 I)$, while $\mathbf{Y}_{true}(t) \sim N_d(\alpha \mathbf{Y}_{true}(t-1), \Sigma_\epsilon(t))$. We choose for $\Sigma_\epsilon(t)$ an exponential spatial covariance function (see Section 5.2.3), so that $\mathbf{Y}_{true}(t) \sim N_d(\alpha \mathbf{Y}_{true}(t-1), \sigma^2 \exp(-h\phi_\epsilon))$.

In this example, we consider a spatial process of size $d = 10$ and $T = 100$ time points, so that \mathbf{Y}_{obs} and \mathbf{Y}_{true} are $T \times d$ matrix. We simulate from a model with parameters: $\Psi = \{\sigma_\xi = \sqrt{0.1}, \alpha = 0.8, \sigma_\epsilon = 0.4, \phi_\epsilon = 0.007\}$.

As mentioned above, estimates are obtained through steps A and B. In particular, to estimate the parameters in step A, we resort to the likelihood resulting from (6.33) and an *InvGamma*(4, 8) prior for σ_ξ^2 . While for the parameters α , σ_ϵ^2 and ϕ_ϵ we resort to the likelihood resulting from (6.34) and priors given by $\alpha \sim Beta(15/4, 2)$, $\sigma_\epsilon^2 \sim InvGamma(20, 14)$, $\phi_\epsilon \sim Gamma(3, 3)$. Finally, for the wrapping coefficients k 's, we choose as prior a Multinomial distribution on values $(-1, 0, 1)$ with related probability vector $(1/4, 1/2, 1/4)$. The choice of this particular prior setting is the result of knowledge gained from Chapter 3 about k 's and from the simulation plan

	True Values	Posterior Median	95% Posterior Interval
σ_ξ	$\sqrt{0.1}$	0.319	[0.297, 0.331]
σ_ϵ	0.4	0.422	[0.367, 0.452]
ϕ_ϵ	0.007	0.0065	[0.0048, 0.0084]
α	0.8	0.842	[0.781, 0.879]

Table 6.1: Parameter estimates of the wrapped spatiotemporal dynamic model.

	Spatiotemporal model
Parameter	Z-score
σ_ξ^2	0.169
σ_ϵ^2	-1.334
ϕ_ϵ	1.154
α	-0.234

Table 6.2: Geweke Z-score diagnostic for the parameters of wrapped spatiotemporal dynamic model.

computed in Section 5.4.1. In these studies we have found that when the complexity of the model increases (and in our case (6.27)-(6.28) is a spatiotemporal dynamic model with quite high number of spatial locations and time points), informative priors helps to improve both the reaching of the MCMC chains convergence and the estimate reliability as well.

To estimate the latent spatial process $Y_{true}(t), t = 1, \dots, T$, instead, the setting in step B regards the initial parameters m_0 and C_0 (see Section 6.3.2) posed equal to zero vector and diagonal unit matrix respectively. The other matrix of the DLM formulation are easily obtained by expressions described in Section 6.3.1. Finally, due to the quite high spatial and temporal dimensions of the model, a large number of chain iterations are necessary to reach the convergence. Here we set a number of iterations equal to 250000 with a burn in of 50000. Moreover we adopt a thin of 10 in order to eliminate the autocorrelation in the chains.

The inference results are reported in Table 6.1 while in Figure 6.1 are depicted the traces of the estimated parameters.

Despite the complexity of the model, the use of informative priors for all parameters and especially for the k 's coefficients produce a very reliable results in terms of estimates and convergence as well. In fact, the posterior interval sizes of the estimates, listed in Table 6.1, are small and the convergence is confirmed by the Geweke diagnostic listed in Table 6.2.

It can be notice that the use of informative priors for Ψ and the introduction of opportune priors (derived from Section 3.1) on k 's, leads to a considerable improving of inference results in terms of estimate reliability. Without this contribute, the computation of complicated model as the spatial model or spatiotemporal model remains intractable.

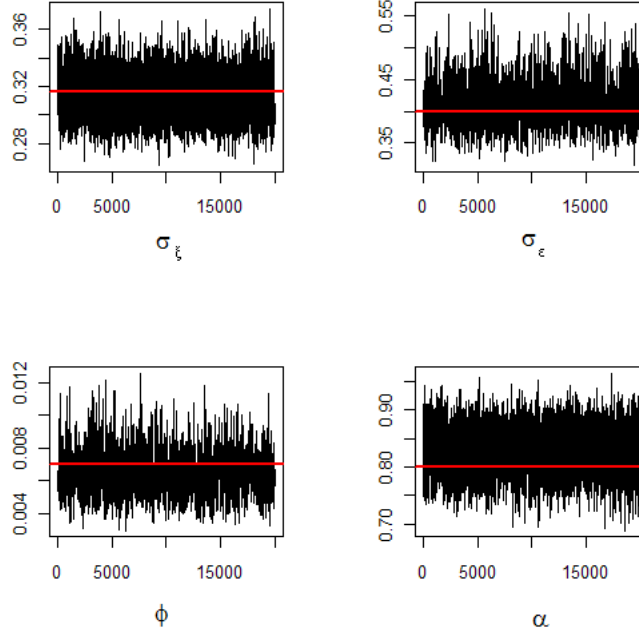


Figure 6.1: Traces of the estimate parameters of wrapped spatiotemporal dynamic model.

6.5 Discussion

In this chapter we give a general overview on spatiotemporal models considering several space-time settings. In particular we focus on models that provide:

- i) additive temporal and spatial effects;
- ii) temporal evolution over sites;
- iii) spatial evolution over time.

For each of these formulations, we provide definitions and Bayesian model specifications.

Next, we introduce the dynamic spatiotemporal modeling (also called state-space modeling), considered a very flexible representation for spatiotemporal structure. In this context we show a very general and fashionable formulation originally provided by Gelfand *et al.* (2005).

Then we derive the wrapped spatiotemporal model as an extension of a wrapped spatial model and a particular case of a measurement error model. Concludes the chapter a simulated example of dynamic wrapped spatiotemporal model. Details about simulation and inference results are provided with particular attention to the computational aspects.

Finally, it is worth to note that the framework of the simulated example can be viewed as a part of a more general calibrating model. To see this, recall the motivating example mentioned in the first chapter in which the target is the calibration

of the WAM (estimated) data through RON data (observed). A general structure for a calibrating model can be the following one:

$$\begin{aligned} X_{WAM}(\mathbf{s}, t) &= f(X_{true}(\mathbf{s}, t)) + \nu_{WAM}(\mathbf{s}, t) \quad \mathbf{s} = s_1, \dots, s_d \\ X_{obs}(\mathbf{s}, t) &= f(X_{true}(\mathbf{s}, t)) + \nu_{obs}(\mathbf{s}, t) \quad t = 1, \dots, T \end{aligned}$$

where the calibration is obtained through the common latent spatiotemporal process $X_{true}(\mathbf{s}, t)$ expressed in both equations of model. The model illustrated in the simulated example basically constitutes a hierarchical development of each of two equations of the above given calibrating model. Thus, we can affirm that the good results obtained in the simulated example, constitute the basis for a future work dealing with the calibrating model computation.

Conclusions and Future Works

The attempt to overcome the difficulty of treating directly wave direction data Bruschi *et al.* (2005) and the research of a method to extend the in line inference procedures to circular data are the starting motives of this thesis.

For circular data, in fact, the usual descriptive and inferential statistical methods cannot be used because of the circular geometry of the sample space. Moreover, the difficulty in managing these data is basically the reason why only “standard” simple models, as linear regression or easy models for temporal time series, have been proposed. In this work, we have developed and extended an approach that can be applied also to structurally and computationally more complicated models as the spatial or spatiotemporal ones.

In Chapter 1 and 2 we have given a fairly comprehensive account of circular data, describing the descriptive and probabilistic aspects. Particular attention it has been given to probabilistic approaches to model circular data through which the probabilistic distributions for circular data are derived. Then, we have focused on the wrapping approach. This method to handle circular data has been already adopted in earlier works as in Coles (1998) and Ravindran (2002) via data augmentation approach. In the former, a bivariate wrapped Normal distribution and a wrapped autoregressive model have been computed. Ravindran, instead, has provided a list of circular distributions derived from wrapping approach together with a linear regression and an autoregressive model. It is worth to note that, notwithstanding the flexibility and the ease of applicability and interpretability of this approach, there are not in literature other examples of the use of the wrapping approach for circular data. In fact, the large part of the works and results about circular data are carried out via intrinsic approach by the von Mises distribution. For this reason, we have decided to deeply analyze the features, advantages and drawbacks of the wrapping approach in Chapter 3. From this study, it has been evidenced that, despite the advantages to have many (infinite) wrapped distributions (as many as the in line distribution are) and the ease of interpretability, the main drawbacks are related to the high number of parameters to estimate (for each circular observation x_j , the relative wrapping coefficient k_j has to be estimated) and the identifiability problem inherent to the wrapping coefficients. In particular, we have realized and demonstrated that the identifiability problem occurs when the data arise large variability. In fact, due to the definition of wrapped distribution itself, the large data variance is confounded with a large value of the wrapping coefficients avoiding to obtain good inference results. Probably, the reason of the limited use of the wrapping approach for circular data regards just these drawbacks.

The main contribution of this thesis deals with the overcoming of these disadvantages. In particular we have demonstrated that the wrapping coefficient values (k -values) that characterized the wrapped Normal distribution are essentially three $(-1, 0, 1)$, and the inference can be limited only upon these values. In other words, we have shown that the wrapped Normal density is well approximated by only three k -values $(-1, 0, 1)$, and it is substantially a mixture of three Normal (in line) distribution characterized by k -values equal to $(-1, 0, 1)$. This knowledge is directly expressed in the inference procedure through the prior on the k 's themselves, solving the identifiability problem and leading to a considerable improvement in the estimate results.

Thus, the Bayesian inference procedure derived in Section 3.4, as a consequence of the study carried out on the wrapping approach, constitutes a crucial result that allows to solve the identifiability problem and, thus, substantially to overcome the peculiar difficulties related to wrapped circular distributions as well as to apply the standard in line models to circular data.

Another key feature of the wrapping approach regards the ease of extending the wrapped distributions to the multivariate case, as illustrated in Chapter 3. Because of the growing amount of assimilated data, the multivariate modeling is becoming a main target of the modern methodologies that often have to manage large data sets that contain different sources of data. In this sense, the von Mises distribution shows all its limits. In fact, its extension to the multivariate case is still an open problem and, even only in the three dimension case, the definition and the parameter estimation of the distribution probability is a challenge (see Mardia *et al.*, 2008). Conversely, the multivariate wrapped Normal distribution presents evident theoretical and practice advantages in that all its marginals are wrapped Normal.

In order to illustrate the ease of applying the wrapped circular distribution to fairly complicated contexts, in Chapter 4 we have given a general overview on measurement error (ME) model needed to derive the circular wrapped version of the ME model. In particular, we have shown that using the wrapped probability distribution properties described in Section 2.4.3 we are able to derive in a intuitive way the wrapped circular ME model. The spatial and spatiotemporal applications in Chapter 5 and 6 constitute two examples of ME models in spatial and spatiotemporal contexts respectively. In particular, in Chapter 5 we have derived the wrapped spatial Normal model as an extension of a multivariate wrapped Normal ME model providing several simulated examples as well. Moreover, in order to appreciate the wrapped multivariate modeling, a simulation plan to understand the behavior of the MCMC estimation algorithm has been projected. From the evidences of the plan, it has been possible to set opportunely the MCMC algorithm for the multivariate high dimensional real data example obtaining, even in this case, good estimate results. Furthermore, it is important to point out that the definition of a spatial structure on a circular distribution different from the wrapped Normal, for example on a von Mises distribution, can be very difficult. Due to the not easy interpretability of the von Mises parameters, it can be very complicated to define, in a meaningful way, a spatial structure.

Perhaps the flexibility of the wrapping approach is even more highlighted and ap-

preciated in the simulated example of Chapter 6. There, after having illustrated the in line spatiotemporal modeling and having derived the corresponding wrapped spatiotemporal model, we have provided a simulated wrapped dynamic spatiotemporal ME model. Notwithstanding the complexity of the in line model, we was able to carry out the corresponding wrapped version and to obtain good estimation results. It is worth to note that the spatial and dynamic spatiotemporal models are originally presented here for the first time.

With regard to future works, the development of the wrapped dynamic spatiotemporal ME model that concludes the Chapter 6 is particularly interesting. In Section 6.5, in fact, we have pointed out as this simulated example substantially constitutes the main structure for the development of a calibrating circular model for the mean wave directions provided by RON and WAM sources (see Section 1.1). More in general, the inferential improvement of the wrapping approach introduced here is a starting point to extend to circular data any in line model in the class of Normal error model, and at the same time, can constitute the basic idea for a future work of extending this approach to any other class of non Normal model. Besides the applications that can derive directly from the extension and generalization of the simulated examples computed in this work, the main contribution carried out in this thesis involves new future work prospects of both computational and data integration aspects.

About the computational aspect, we believe that more efficient computing algorithms can be used for our purpose. Recent MCMC developments regard adaptive algorithms to automatically tune the Markov chain parameters during a run (Haario *et al.*, 2005). These new methods could provide a very useful tool to improve the MCMC performance of our estimation procedure, especially in high dimensional multivariate contexts. Moreover, the approximation of the wrapped Normal density by a mixture of three in line Normal densities depending on three k 's values $(-1, 0, 1)$ suggests the non Bayesian inferential method based on the EM algorithm. With this approximation the wrapped Normal parameter estimation via EM wouldn't ought to present particular difficulties. Moreover, it can be interesting to compare this last EM estimation procedure with the EM estimation results carried out by Fisher and Lee (1994). Even in this work, the k coefficients were treated as missing data but the limitation of that procedure was its complexity: only low parameter dimension wrapped models have been estimated. Thus, the comparison between the Fisher and Lee's EM procedure and the EM that exploits the mixture of the three Normal densities can be useful to asses the possible computation improvement provided by the latter algorithm.

Regarding the data integration aspect, we point out that it is strictly related to the recent growing amount of data that involves the management of large data sets of different sources. Thus, the need to develop methods that allow to integrate circular and in line data in mixed models appears a focus work. Considering, for example, the regression model, we find some applications of linear response with circular variables in Mardia (1976) and in Johnson and Wehrly (1978). A more challenge application regards regression of circular variable on in line explanatory variables. A first solution based on the intrinsic approach by von Mises distribution

has been proposed in Gould (1969) and Laycock (1975), where the in line variables are mapped into circle by the *link function* $y \mapsto \beta y \pmod{2\pi}$, i.e. adopting a solution substantially equivalent to the wrapping approach. Successively, Fisher and Lee (1992) proposed to use a one-to-one link function. Fisher (1993, Sec. 6.4) shows an iterative method to calculate the maximum likelihood estimates under opportune choice of the link function. But, even in this last case, the use of the intrinsic approach is limited from both computational problems, that arise especially with large number of explanatory variables, and the choice of the link function that influences the inference results.

Thus, even for mixed model, the wrapping approach appears preferable to the intrinsic approach and, conversely of the latter, offers an easier interpretation of the parameters.

Appendix A

Simulation plan

Simul identif	Transition density parameters			Prior on μ	Prior on ϕ (true=0.05)		Prior on σ^2 (true=1)	μ true values	Variance of μ prior sigma-mu out-dia		iter.
	a	b	c								
d6-1	1	1.3	2	MVN(0,sigma.mu)	gamma(3,40)		invgamma(8,9)	(-0.7,-0.5,-0.2,0.3,0.5,0.7)	runif(0.5,1)	runif(0.5,1.6)	15000
d6-2	1	1.3	2	MVN(0,sigma.mu)	gamma(3,40)		invgamma(8,9)	(-0.7,-0.5,-0.2,0.3,0.5,0.7)	runif(0.8,2)	runif(0.8,2.4)	15000
d6-3	1.8	1.3	2	MVN(0,sigma.mu)	gamma(3,40)		invgamma(8,9)	(-2.2,-1.8,-1.2,0.7,1.4,2)	runif(1.2,2)	runif(1.3,2)	15000
d6-4	1	1.3	2	MVN(0,sigma.mu)	gamma(3,40)		invgamma(8,9)	(-2.2,-1.8,-1.2,0.7,1.4,2)	runif(1.2,5)	runif(1.3,5)	15000
d6-5	1	1.3	2	MVN((-0.7,-0.5,-0.2,0.3,0.5,0.7),sigma.mu)	gamma(3,40)		invgamma(8,9)	(-0.7,-0.5,-0.2,0.3,0.5,0.7)	runif(0.5,1)	runif(0.5,1.6)	15000
d6-6	1	1.3	2	MVN((-0.7,-0.5,-0.2,0.3,0.5,0.7),sigma.mu)	gamma(3,40)		invgamma(8,9)	(-0.7,-0.5,-0.2,0.3,0.5,0.7)	runif(0.8,2)	runif(0.8,2.4)	15000
d6-7	1.8	1.3	2	MVN((-2.2,-1.8,-1.2,0.7,1.4,2),sigma.mu)	gamma(3,40)		invgamma(8,9)	(-2.2,-1.8,-1.2,0.7,1.4,2)	runif(1.2,2)	runif(1.3,2)	15000
d6-8	1.8	1.3	2	MVN((-2.2,-1.8,-1.2,0.7,1.4,2),sigma.mu)	gamma(3,40)		invgamma(8,9)	(-2.2,-1.8,-1.2,0.7,1.4,2)	runif(1.2,5)	runif(1.3,5)	15000
d6-9	1	1.3	2	MVN(0,sigma.mu)	gamma(2,20)		invgamma(8,9)	(-0.7,-0.5,-0.2,0.3,0.5,0.7)	runif(0.5,1)	runif(0.5,1.6)	15000
d6-10	1	1.3	2	MVN(0,sigma.mu)	gamma(2,20)		invgamma(8,9)	(-0.7,-0.5,-0.2,0.3,0.5,0.7)	runif(0.8,2)	runif(0.8,2.4)	15000
d6-11	1.8	1.3	2	MVN(0,sigma.mu)	gamma(2,20)		invgamma(8,9)	(-2.2,-1.8,-1.2,0.7,1.4,2)	runif(1.2,2)	runif(1.3,2)	15000
d6-12	1.8	1.3	2	MVN(0,sigma.mu)	gamma(2,20)		invgamma(8,9)	(-2.2,-1.8,-1.2,0.7,1.4,2)	runif(1.2,5)	runif(1.3,5)	15000
d6-13	1	1.3	2	MVN((-0.7,-0.5,-0.2,0.3,0.5,0.7),sigma.mu)	gamma(2,20)		invgamma(8,9)	(-0.7,-0.5,-0.2,0.3,0.5,0.7)	runif(0.5,1)	runif(0.5,1.6)	15000
d6-14	1	1.3	2	MVN((-0.7,-0.5,-0.2,0.3,0.5,0.7),sigma.mu)	gamma(2,20)		invgamma(8,9)	(-0.7,-0.5,-0.2,0.3,0.5,0.7)	runif(0.8,2)	runif(0.8,2.4)	15000
d6-15	1.8	1.3	2	MVN((-2.2,-1.8,-1.2,0.7,1.4,2),sigma.mu)	gamma(2,20)		invgamma(8,9)	(-2.2,-1.8,-1.2,0.7,1.4,2)	runif(1.2,2)	runif(1.3,2)	15000
d6-16	1.8	1.3	2	MVN((-2.2,-1.8,-1.2,0.7,1.4,2),sigma.mu)	gamma(2,20)		invgamma(8,9)	(-2.2,-1.8,-1.2,0.7,1.4,2)	runif(1.2,5)	runif(1.3,5)	15000
d6-17	1	1.3	2	MVN(0,sigma.mu)	gamma(3,40)		invgamma(4,8.5)	(-0.7,-0.5,-0.2,0.3,0.5,0.7)	runif(0.5,1)	runif(0.5,1.6)	15000
d6-18	1	1.3	2	MVN(0,sigma.mu)	gamma(3,40)		invgamma(4,8.5)	(-0.7,-0.5,-0.2,0.3,0.5,0.7)	runif(0.8,2)	runif(0.8,2.4)	15000
d6-19	1.8	1.3	2	MVN(0,sigma.mu)	gamma(3,40)		invgamma(4,8.5)	(-2.2,-1.8,-1.2,0.7,1.4,2)	runif(1.2,2)	runif(1.3,2)	15000
d6-20	1.8	1.3	2	MVN(0,sigma.mu)	gamma(3,40)		invgamma(4,8.5)	(-2.2,-1.8,-1.2,0.7,1.4,2)	runif(1.2,5)	runif(1.3,5)	15000
d6-21	1	1.3	2	MVN((-0.7,-0.5,-0.2,0.3,0.5,0.7),sigma.mu)	gamma(3,40)		invgamma(4,8.5)	(-0.7,-0.5,-0.2,0.3,0.5,0.7)	runif(0.5,1)	runif(0.5,1.6)	15000
d6-22	1	1.3	2	MVN((-0.7,-0.5,-0.2,0.3,0.5,0.7),sigma.mu)	gamma(3,40)		invgamma(4,8.5)	(-0.7,-0.5,-0.2,0.3,0.5,0.7)	runif(0.8,2)	runif(0.8,2.4)	15000
d6-23	1.8	1.3	2	MVN((-2.2,-1.8,-1.2,0.7,1.4,2),sigma.mu)	gamma(3,40)		invgamma(4,8.5)	(-2.2,-1.8,-1.2,0.7,1.4,2)	runif(1.2,2)	runif(1.3,2)	15000
d6-24	1.8	1.3	2	MVN((-2.2,-1.8,-1.2,0.7,1.4,2),sigma.mu)	gamma(3,40)		invgamma(4,8.5)	(-2.2,-1.8,-1.2,0.7,1.4,2)	runif(1.2,5)	runif(1.3,5)	15000

Table A.1: Simulation plan scheme for wrapped spatial process with dimension $d = 6$

Simul identif	Transition density parameters			Prior on μ	Prior on ϕ (true=0.05)	Prior on σ^2 (true=1)	μ true values	Variance of μ prior		N. iter.
	a	b	c					sigma.mu	out-diag	
d15-1	1	1.3	2	MVN(0,sigma.mu)	gamma(3,40)	invgamma(8,9)	(-0.7,-0.6,...,0.6,0.7)	runif(0.5,1)	runif(0.5,1.6)	100000
d15-2	1	1.3	2	MVN(0,sigma.mu)	gamma(3,40)	invgamma(8,9)	(-0.7,-0.6,...,0.6,0.7)	runif(0.8,2)	runif(0.8,2.4)	100000
d15-3	1.8	1.3	2	MVN(0,sigma.mu)	gamma(3,40)	invgamma(8,9)	(-1.5,-1.3,...,1.6,1.8)	runif(1.2,2)	runif(1.3,2)	100000
d15-4	1.8	1.3	2	MVN(0,sigma.mu)	gamma(3,40)	invgamma(8,9)	(-1.5,-1.3,...,1.6,1.8)	runif(1.2,2)	runif(1.3,5)	100000
d15-5	1	1.3	2	MVN((-0.7,-0.6,...,0.6,0.7),sigma.mu)	gamma(3,40)	invgamma(8,9)	(-0.7,-0.6,...,0.6,0.7)	runif(0.5,1)	runif(0.5,1.6)	100000
d15-6	1	1.3	2	MVN((-0.7,-0.6,...,0.6,0.7),sigma.mu)	gamma(3,40)	invgamma(8,9)	(-0.7,-0.6,...,0.6,0.7)	runif(0.8,2)	runif(0.8,2.4)	100000
d15-7	1.8	1.3	2	MVN((-1.5,-1.3,...,1.6,1.8),sigma.mu)	gamma(3,40)	invgamma(8,9)	(-1.5,-1.3,...,1.6,1.8)	runif(1.2,2)	runif(1.3,2)	100000
d15-8	1.8	1.3	2	MVN((-1.5,-1.3,...,1.6,1.8),sigma.mu)	gamma(3,40)	invgamma(8,9)	(-1.5,-1.3,...,1.6,1.8)	runif(1.2,5)	runif(0.5,1.6)	100000
d15-9	1	1.3	2	MVN(0,sigma.mu)	gamma(2,20)	invgamma(8,9)	(-0.7,-0.6,...,0.6,0.7)	runif(0.5,1)	runif(0.5,1.6)	100000
d15-10	1	1.3	2	MVN(0,sigma.mu)	gamma(2,20)	invgamma(8,9)	(-0.7,-0.6,...,0.6,0.7)	runif(0.8,2)	runif(0.8,2.4)	100000
d15-11	1.8	1.3	2	MVN(0,sigma.mu)	gamma(2,20)	invgamma(8,9)	(-1.5,-1.3,...,1.6,1.8)	runif(1.2,2)	runif(1.3,2)	100000
d15-12	1.8	1.3	2	MVN(0,sigma.mu)	gamma(2,20)	invgamma(8,9)	(-1.5,-1.3,...,1.6,1.8)	runif(1.2,5)	runif(1.3,5)	100000
d15-13	1	1.3	2	MVN((-0.7,-0.6,...,0.6,0.7),sigma.mu)	gamma(2,20)	invgamma(8,9)	(-0.7,-0.6,...,0.6,0.7)	runif(0.5,1)	runif(0.5,1.6)	100000
d15-14	1	1.3	2	MVN((-0.7,-0.6,...,0.6,0.7),sigma.mu)	gamma(2,20)	invgamma(8,9)	(-0.7,-0.6,...,0.6,0.7)	runif(0.8,2)	runif(0.8,2.4)	100000
d15-15	1.8	1.3	2	MVN((-1.5,-1.3,...,1.6,1.8),sigma.mu)	gamma(2,20)	invgamma(8,9)	(-1.5,-1.3,...,1.6,1.8)	runif(1.2,2)	runif(1.3,2)	100000
d15-16	1.8	1.3	2	MVN((-2,-1.2),sigma.mu)	gamma(2,20)	invgamma(8,9)	(-1.5,-1.3,...,1.6,1.8)	runif(1.2,5)	runif(1.3,5)	100000
d15-17	1	1.3	2	MVN(0,sigma.mu)	gamma(3,40)	invgamma(4.8,5)	(-0.7,-0.6,...,0.6,0.7)	runif(0.5,1)	runif(0.5,1.6)	100000
d15-18	1	1.3	2	MVN(0,sigma.mu)	gamma(3,40)	invgamma(4.8,5)	(-0.7,-0.6,...,0.6,0.7)	runif(0.8,2)	runif(0.8,2.4)	100000
d15-19	1.8	1.3	2	MVN(0,sigma.mu)	gamma(3,40)	invgamma(4.8,5)	(-1.5,-1.3,...,1.6,1.8)	runif(1.2,2)	runif(1.3,2)	100000
d15-20	1.8	1.3	2	MVN(0,sigma.mu)	gamma(3,40)	invgamma(4.8,5)	(-1.5,-1.3,...,1.6,1.8)	runif(1.2,5)	runif(1.3,5)	100000
d15-21	1	1.3	2	MVN((-0.7,-0.6,...,0.6,0.7),sigma.mu)	gamma(3,40)	invgamma(4.8,5)	(-0.7,-0.6,...,0.6,0.7)	runif(0.5,1)	runif(0.5,1.6)	100000
d15-22	1	1.3	2	MVN((-0.7,-0.6,...,0.6,0.7),sigma.mu)	gamma(3,40)	invgamma(4.8,5)	(-0.7,-0.6,...,0.6,0.7)	runif(0.8,2)	runif(0.8,2.4)	100000
d15-23	1.8	1.3	2	MVN((-1.5,-1.3,...,1.6,1.8),sigma.mu)	gamma(3,40)	invgamma(4.8,5)	(-1.5,-1.3,...,1.6,1.8)	runif(1.2,2)	runif(1.3,2)	100000
d15-24	1.8	1.3	2	MVN((-1.5,-1.3,...,1.6,1.8),sigma.mu)	gamma(3,40)	invgamma(4.8,5)	(-1.5,-1.3,...,1.6,1.8)	runif(1.2,5)	runif(1.3,5)	100000

Table A.2: Simulation plan scheme for wrapped spatial process with dimension $d = 15$

Parameters	d3-1	d3-2	d3-3	d3-4	d3-5	d3-6	d3-7	d3-8	d3-9	d3-10	d3-11	d3-12
True=0.5; μ_1	-0.2	-0.18	-0.1	-0.2	-0.3	-0.25	-0.25	-0.3	-0.3	-0.3	-0.1	-0.15
95% Posterior Interval	[-1.6, 1.2]	[-1.5, 1.1]	[-1.5, 1.3]	[-1.5, 1.4]	[-1.6, 1.4]	[-1.7, 1.7]	[-1.7, 1.3]	[-1.3, 1.3]	[-1.6, 1.2]	[-1.8, 1.3]	[-1.4, 1.6]	[-1.5, 1.4]
True=0.2; μ_2	-0.04	-0.02	0.1	-0.01	-0.12	-0.1	-0.13	-0.12	-0.14	-0.17	-0.11	-0.07
95% Posterior Interval	[-1.5, -1.3]	[-1.5, 1.2]	[-1.4, -1.4]	[-1.5, 1.3]	[-1.5, -1.3]	[-1.4, 1.4]	[-1.2, -1.5]	[-1.5, 1.2]	[-1.5, -1.3]	[-1.5, 1.2]	[-1.4, -1.4]	[-1.5, 1.3]
True=0.5; μ_3	0.3	0.2	0.18	0.2	0.33	0.28	0.26	0.26	0.31	0.3	0.13	0.16
95% Posterior Interval	[-1.5, 1.8]	[-1.5, 1.7]	[-1.4, 1.5]	[-1.3, 1.5]	[-1.4, 1.6]	[-1.5, 1.8]	[-1.5, 1.7]	[-1.4, 1.5]	[-1.3, 1.5]	[-1.4, 1.6]	[-1.4, 1.5]	[-1.3, 1.5]
True=1; $\hat{\sigma}^2$	1.01	1.01	1.2	1.1	1.02	1.04	1.07	1.1	1.15	0.98	0.96	1.02
95% Posterior Interval	[0.71, 1.4]	[0.73, 1.35]	[0.73, 1.5]	[0.70, 1.52]	[0.69, 1.35]	[0.81, 1.5]	[0.71, 1.4]	[0.73, 1.35]	[0.73, 1.5]	[0.70, 1.52]	[0.69, 1.35]	[0.81, 1.5]
True=0.05; $\hat{\phi}$	0.04	0.05	0.08	0.06	0.04	0.03	0.06	0.07	0.05	0.06	0.08	0.03
95% Posterior Interval	[0.005, 0.14]	[0.005, 0.13]	[0.006, 0.15]	[0.004, 0.16]	[0.002, 0.14]	[0.001, 0.14]	[0.005, 0.13]	[0.006, 0.15]	[0.004, 0.16]	[0.002, 0.14]	[0.006, 0.14]	[0.003, 0.13]

Table A.3: Simulation plan results for wrapped spatial process with dimension $d = 3$ (first part)

Parameters	d3-13	d3-14	d3-15	d3-16	d3-17	d3-18	d3-19	d3-20	d3-21	d3-22	d3-23	d3-24
True=-0.5; μ_1	-0.2	-0.22	-0.26	-0.28	-0.3	-0.32	-0.26	-0.27	-0.4	-0.39	-0.35	-0.34
95% Posterior Interval	[-1.6, 1.2]	[-1.5, 1.3]	[-1.4, 1.4]	[-1.5, 1.5]	[-1.6, 1.2]	[-1.5, 1.4]	[-1.7, 1.5]	[-1.4, 1.7]	[-1.5, 1.5]	[-1.6, 1.2]	[-1.5, 1.2]	[-1.6, 1.3]
True=-0.2; μ_2	-0.1	-0.11	-0.08	-0.09	-0.13	-0.13	-0.08	-0.07	-0.17	-0.16	-0.15	-0.12
95% Posterior Interval	[-1.5, -1.3]	[-1.6, 1.4]	[-1.2, -1.5]	[-1.5, 1.2]	[-1.5, -1.3]	[-1.5, -1.3]	[-1.6, 1.4]	[-1.2, -1.5]	[-1.5, 1.2]	[-1.5, -1.3]	[-1.5, -1.3]	[-1.5, 1.2]
True=0.5; μ_3	0.25	0.28	0.26	0.28	0.32	0.31	0.29	0.3	0.45	0.4	0.39	0.38
95% Posterior Interval	[-1.4, 1.6]	[-1.5, 1.8]	[-1.5, 1.7]	[-1.4, 1.5]	[-1.5, 1.8]	[-1.5, 1.7]	[-1.4, 1.5]	[-1.3, 1.5]	[-1.4, 1.6]	[-1.4, 1.5]	[-1.3, 1.5]	[-1.4, 1.6]
True=1; σ^2	1.01	0.89	0.88	1.04	1.02	1.03	1.04	1.03	0.99	0.97	1.02	1.1
95% Posterior Interval	[0.71, 1.4]	[0.81, 1.5]	[0.71, 1.4]	[0.73, 1.35]	[0.73, 1.5]	[0.73, 1.5]	[0.70, 1.52]	[0.69, 1.35]	[0.81, 1.5]	[0.71, 1.4]	[0.81, 1.5]	[0.71, 1.4]
True=0.05; ϕ	0.05	0.04	0.07	0.06	0.04	0.04	0.06	0.07	0.05	0.05	0.06	0.04
95% Posterior Interval	[0.004, 0.16]	[0.002, 0.14]	[0.001, 0.14]	[0.005, 0.13]	[0.003, 0.15]	[0.004, 0.16]	[0.002, 0.14]	[0.006, 0.14]	[0.004, 0.16]	[0.002, 0.14]	[0.005, 0.14]	[0.005, 0.13]

Table A.4: Simulation plan results for wrapped spatial process with dimension $d = 3$ (second part)

Parameters	d3-1	d3-2	d3-3	d3-4	d3-5	d3-6	d3-7	d3-8	d3-9	d3-10	d3-11	d3-12
True= $0.7; \mu_1$	-0.4	-0.36	-0.35	-0.38	-0.41	-0.37	-0.36	-0.39	-0.42	-0.4	-0.32	-0.3
95% Posterior Interval	[-2.01, 2.13]	[-1.91, 1.93]	[-1.89, 1.93]	[-2.01, 2.03]	[-2.11, 2.15]	[-1.98, 2.03]	[-2.01, 2.13]	[-2.01, 1.99]	[-2.09, 2.16]	[-2.11, 2.17]	[-1.98, 1.99]	[-1.88, 1.89]
True= $0.5; \mu_1$	-0.2	-0.18	-0.1	-0.2	-0.3	-0.25	-0.25	-0.3	-0.3	-0.3	-0.1	-0.15
95% Posterior Interval	[-1.6, 1.2]	[-1.5, 1.1]	[-1.5, 1.3]	[-1.5, 1.4]	[-1.6, 1.4]	[-1.7, 1.7]	[-1.7, 1.3]	[-1.3, 1.3]	[-1.6, 1.2]	[-1.8, 1.3]	[-1.4, 1.6]	[-1.5, 1.4]
True= $0.2; \mu_2$	-0.04	-0.02	0.1	-0.01	-0.12	-0.1	-0.13	-0.12	-0.14	-0.17	-0.11	-0.07
95% Posterior Interval	[-1.5, -1.3]	[-1.5, 1.2]	[-1.4, -1.4]	[-1.5, 1.3]	[-1.5, -1.3]	[-1.6, 1.4]	[-1.2, -1.5]	[-1.5, 1.2]	[-1.5, -1.3]	[-1.5, 1.2]	[-1.4, -1.4]	[-1.5, 1.3]
True= $0.3; \mu_2$	-0.12	0.18	0.15	0.18	0.21	0.19	0.18	0.2	0.22	0.23	0.17	0.16
95% Posterior Interval	[-1.9, 1.5]	[-1.8, 1.5]	[-1.9, 1.6]	[-1.7, 1.6]	[-1.8, 1.6]	[-1.7, 1.5]	[-1.5, 1.5]	[-1.4, 1.6]	[-1.4, 1.7]	[-1.8, 1.6]	[-1.9, 1.5]	[-1.9, 1.8]
True= $0.5; \mu_3$	0.3	0.2	0.18	0.2	0.33	0.28	0.26	0.26	0.31	0.3	0.13	0.16
95% Posterior Interval	[-1.5, 1.8]	[-1.5, 1.7]	[-1.4, 1.5]	[-1.3, 1.5]	[-1.4, 1.6]	[-1.5, 1.8]	[-1.5, 1.7]	[-1.4, 1.5]	[-1.3, 1.5]	[-1.4, 1.6]	[-1.4, 1.5]	[-1.3, 1.5]
True= $0.7; \mu_1$	0.38	0.39	0.36	0.36	0.43	0.39	0.37	0.41	0.43	0.45	0.37	0.32
95% Posterior Interval	[-1.2, 2.24]	[-1.3, 2.3]	[-1.3, 2.4]	[-1.2, 2.4]	[-1.3, 2.2]	[-1.4, 2.4]	[-1.5, 2.5]	[-1.1, 2.6]	[-1.1, 2.5]	[-1.2, 2.6]	[-1.2, 2.24]	[-1.2, 2.24]
True= $1; \hat{\sigma}^2$	1.01	1.01	1.2	1.1	1.02	1.04	1.07	1.1	1.15	0.98	0.96	1.02
95% Posterior Interval	[0.71, 1.4]	[0.73, 1.35]	[0.73, 1.5]	[0.70, 1.52]	[0.69, 1.35]	[0.81, 1.5]	[0.71, 1.4]	[0.73, 1.35]	[0.73, 1.5]	[0.70, 1.52]	[0.69, 1.35]	[0.81, 1.5]
True= $0.05; \phi$	0.04	0.05	0.08	0.06	0.04	0.03	0.06	0.07	0.05	0.06	0.08	0.03
95% Posterior Interval	[0.005, 0.14]	[0.005, 0.13]	[0.006, 0.15]	[0.004, 0.16]	[0.002, 0.14]	[0.001, 0.14]	[0.005, 0.13]	[0.006, 0.15]	[0.004, 0.16]	[0.002, 0.14]	[0.006, 0.14]	[0.003, 0.13]

Table A.5: Simulation plan results for wrapped spatial process with dimension $d = 6$ (first part)

Parameters	d3-13	d3-14	d3-15	d3-16	d3-17	d3-18	d3-19	d3-20	d3-21	d3-22	d3-23	d3-24
True=-0.7; μ_1	-0.33	-0.32	-0.36	-0.38	-0.35	-0.38	-0.31	-0.37	-0.49	-0.48	-0.45	-0.41
95% Posterior Interval	[-1.89, 1.93]	[-2.01, 2.03]	[-2.11, 2.15]	[-1.91, 1.93]	[-1.89, 1.93]	[-2.01, 2.03]	[-1.98, 2.03]	[-2.01, 2.13]	[-2.01, 1.99]	[-2.11, 2.15]	[-1.91, 1.93]	[-2.01, 2.03]
True=-0.5; μ_1	-0.2	-0.22	-0.26	-0.28	-0.3	-0.32	-0.26	-0.27	-0.4	-0.39	-0.35	-0.34
95% Posterior Interval	[-1.6, 1.2]	[-1.5, 1.3]	[-1.4, 1.4]	[-1.5, 1.5]	[-1.6, 1.2]	[-1.5, 1.4]	[-1.7, 1.5]	[-1.4, 1.7]	[-1.5, 1.5]	[-1.6, 1.2]	[-1.5, 1.2]	[-1.6, 1.3]
True=-0.2; μ_2	-0.1	-0.11	-0.08	-0.09	-0.13	-0.13	-0.08	-0.07	-0.17	-0.16	-0.15	-0.12
95% Posterior Interval	[-1.5, -1.3]	[-1.6, 1.4]	[-1.2, -1.5]	[-1.5, 1.2]	[-1.5, -1.3]	[-1.5, -1.3]	[-1.6, 1.4]	[-1.2, -1.5]	[-1.5, 1.2]	[-1.5, -1.3]	[-1.5, -1.3]	[-1.5, 1.2]
True=0.3; μ_2	0.19	0.18	0.22	0.12	0.26	0.21	0.23	0.28	0.22	0.24	0.22	0.23
95% Posterior Interval	[-1.7, 1.5]	[-1.5, 1.5]	[-1.4, 1.7]	[-1.9, 1.8]	[-1.4, 1.5]	[-1.3, 1.5]	[-1.4, 1.6]	[-1.5, 1.8]	[-1.5, 1.9]	[-1.6, 1.9]	[-1.4, 1.9]	[-1.4, 1.8]
True=0.5; μ_3	0.25	0.28	0.26	0.28	0.32	0.31	0.29	0.3	0.45	0.4	0.39	0.38
95% Posterior Interval	[-1.4, 1.6]	[-1.5, 1.8]	[-1.5, 1.7]	[-1.4, 1.5]	[-1.5, 1.8]	[-1.5, 1.7]	[-1.4, 1.5]	[-1.3, 1.5]	[-1.4, 1.6]	[-1.4, 1.5]	[-1.3, 1.5]	[-1.4, 1.6]
True=0.7; μ_1	0.37	0.36	0.36	0.33	0.39	0.4	0.32	0.33	0.5	0.51	0.36	0.37
95% Posterior Interval	[-1.3, 2.4]	[-1.2, 2.4]	[-1.3, 2.2]	[-1.4, 2.4]	[-1.2, 2.24]	[-1.3, 2.3]	[-1.1, 2.6]	[-1.1, 2.5]	[-1.2, 2.6]	[-1.2, 2.4]	[-1.3, 2.2]	[-1.4, 2.4]
True=1; ϕ^2	1.01	0.89	0.88	1.04	1.02	1.03	1.04	1.03	0.99	0.97	1.02	1.1
95% Posterior Interval	[0.71, 1.4]	[0.81, 1.5]	[0.71, 1.4]	[0.73, 1.35]	[0.73, 1.5]	[0.73, 1.5]	[0.70, 1.52]	[0.69, 1.35]	[0.81, 1.5]	[0.71, 1.4]	[0.81, 1.5]	[0.71, 1.4]
True=0.05; ϕ	0.05	0.04	0.07	0.06	0.04	0.04	0.06	0.07	0.05	0.05	0.06	0.04
95% Posterior Interval	[0.004, 0.16]	[0.002, 0.14]	[0.001, 0.14]	[0.005, 0.13]	[0.003, 0.15]	[0.004, 0.16]	[0.002, 0.14]	[0.006, 0.14]	[0.004, 0.16]	[0.002, 0.14]	[0.005, 0.14]	[0.005, 0.13]

Table A.6: Simulation plan results for wrapped spatial process with dimension $d = 6$ (second part)

Bibliography

- Adcock, J. R. (1877). Note on the method of least squares. *Analyst*, **4**, 183–184.
- Arsham, H. (1988). Kuiper’s p-value as a measuring tool and decision procedure for the goodness-of-fit test. *Journal of Applied Statistics*, **15**, 131–135.
- Banerjee, S., Bradley, P. C., and Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall/CRC.
- Barbieri, M. M. (1996). *Metodi MCMC nell’Inferenza Statistica*. Società Italiana di Statistica -Monografie-.
- Batschelet, E. (1965). *Statistical Methods for the Analysis of Problems in Animal Orientation and Certain Biological Rhythms*. Amer. Inst. Biol. Sci., Washington.
- Berkson, J. (1950). Are there two regressions? *Journal of American Statistical Association*, **45**, 164–180.
- Berry, S. M., Carroll, R. J., and Ruppert, D. (2002). Bayesian smoothing and regression splines for measurement error problem. *Journal of American Statistical Association*, **97**, 160–169.
- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, Ser. B*, **36**, 192–236.
- Besag, J., Green, P., Higdon, D., and Mengersen, K. (1995). Bayesian computation and stochastic systems. *Statistical Science*, **10**, 3–41.
- Blake, A. and Marinos, C. (1990). Shape from texture-estimation, isotropy and moments. *Artificial Intelligence*, **45**, 323–380.
- Bowden, R. J. and Turkington, D. A. (1984). *Instrumental Variables*. Cambridge University Press, Cambridge.
- Breckling, J. (1989). *The Analysis of Directional Time Series: Applications to Wind Speed and Direction*. Springer-Verlag.
- Bruschi, A., Conti, P. L., Corsini, S., Divino, F., Ferrari, C., Inghilesi, R., Jona Lasinio, G., Leone, M., and Orasi, A. (2005). Assimilation of meteo-marine data in a coastal forecasting system:case study on the adriatic sea. *Proceeding Workshop Archimede, APAT*.

- Carroll, R. J., Ruppert, D., and Stefanski, L. A. (1995). *Measurement Error in Nonlinear Models*. Chapman and Hall, London.
- Case, K. E. and Shiller, R. J. (1989). The efficiency of the market for single family homes. *American Economic Review*, **79**, 125–137.
- Cheng, C. and Van Ness, J. W. (1999). *Statistical Regression with Measurement Error*. Arnold.
- Cheng, C.-L. (1994). On generalized least squares and least squares. *Unpublished manuscript. Institute of Statistical Sciences, Academia Sinica, Taiwan, Rep. of China*.
- Chib, S. and Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika*, **85**, 347–361.
- Coles, S. (1998). Inference for circular distributions and processes. *Statistics and Computing*, **8**, 105–113.
- Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. Chapman and Hall, London.
- Cressie, N. A. C. (1993). *Statistical for Spatial Data*. 2nd ed. Wiley, New York.
- Damien, P. and Walker, S. (1999). A full bayesian analysis of circular data using von mises distribution. *The Canadian Journal of Statistics*, **27**, 291–298.
- Daniel, M. J. and Gatsonis, C. (1999). Hierarchical generalized linear models in the analysis of deviation in health care utilization. *Journal of American Statistical Association*, **94**, 29–42.
- Diggle, P. J., Tawn, J. A., and Moyeed, R. A. (1998). Model-based geostatistics (with discussion). *Journal of the Royal Statistical Society, Ser. C*, **47**, 299–350.
- Durbin, J. (1973). *Distribution Theory for Tests Based on the Sample Distribution Function*. Society for Industrial and Applied Mathematics, Philadelphia.
- Ecker, M. D. and Gelfand, A. E. (1997). Bayesian variogram modeling for an isotropic spatial process. *Journal of Agricultural, Biological and Environmental Statistics*, **2**, 347–369.
- Fisher, N. I. (1993). *Statistical Analysis of Circular Data*. Cambridge.
- Fisher, N. I. and Lee, A. J. (1992). Regression models for an angular response. *Biometrics*, **48**, 665–677.
- Fisher, N. I. and Lee, A. J. (1994). Time series analysis of circular data. *Journal of the Royal Statistical Society, Ser. B*, **56**, 327–339.
- Fuller, W. A. (1987). *Measurement Error Models*. John Wiley and Sons, New York.

- Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of American Statistical Association*, **85**, 398–409.
- Gelfand, A. E., Sahu, S. K., and Carlin, B. P. (1995). Efficient parameterizations for generalized linear mixed models. *Bayesian Statistics*, **5**, 165–180.
- Gelfand, A. E., Ecker, M. D., Knight, J. R., and F., S. C. (2004). The dynamics of location in home price. *The Journal of Real Estate Finance and Economics*, **29**, 149–166.
- Gelfand, A. E., Banerjee, S., and Gamerman, D. (2005). Spatial process modelling for univariate and multivariate dynamic spatial data. *Environmetrics*, **16**, 465–479.
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian analysis*, **1**, 515–534.
- Gelman, A., Roberts, G. O., and Gilks, W. R. (1996). Efficient metropolis jumping rules. *Bayesian Statistics*, **5**, 599–607.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian Data Analysis*. Chapman and Hall/ CRC Press.
- Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. *Bayesian Statistics*, **4**, 169–193.
- Ghosh, P., Bayes, C. L., and Lachos, V. H. (2008). A robust bayesian approach to null intercept measurement error model with application to dental data. *Computational Statistics and Data Analysis*.
- Gould, A. L. (1969). A regression technique for angular variates. *Biometrics*, **25**, 683–700.
- Guttorp, P. and Lockhart, R. A. (1988). Finding the location of a signal - a bayesian analysis. *Journal of American Statistical Association*, **83**, 322–330.
- Haario, H., Saksman, E., and Tamminen, J. (2005). Componentwise adaptation for high dimensional mcmc. *Computational Statistics*, **20**, 265–273.
- Harrison, D. and Kanji, G. K. (1988). The development of analysis of variance for circular data. *Journal of Applied Statistics*, **15**, 197–224.
- Harville, D. A. and Zimmerman, A. G. (1996). The posterior distribution of the fixed and random effect in a mixed-effect linear model. *Journal of Statistical Computation and Simulation*, **54**, 211–229.
- Hawkins, D. M. and Cressie, N. (1984). Robust kriging -a proposal. *Journal of the International Association for Mathematical Geology*, **16**, 3–18.

- Hill, R. C., Sirmans, C. F., and Knight, J. R. (1999). A random walk down main street. *Regional Science and Urban Economics*, **29**, 89–103.
- Hughes, G. (2007). *Multivariate and Time Series Models for Circular Data with Applications to Protein Conformational Angles*. Ph.D. thesis, University of Leeds, Leeds, England, UK.
- Ishwarant, H. and Gatsonis, C. A. (2000). A general class of hierarchical ordinal regression models with application to correlated roc analysis. *The Canadian Journal of Statistics*, **28**, 731–750.
- Jammalamadaka, S. R. and SenGupta, A. (2001). *Topics in Circular Statistics*. World Scientific.
- Jeffreys, H. (1948). *Theory of Probability, 2nd ed.* Oxford Univ. Press.
- Johnson, R. A. and Wehrly, T. (1978). Some angular-linear distributions and related regression models. *Journal of American Statistical Association*, **73**, 602–606.
- Karson, M. J., Linder, E., and Sinha, D. (1999). Bayesian analysis and computation for spatial prediction (with discussion). *Environmental and Ecological Statistics*, **6**, 147–182.
- Kass, R. E. and Natarajan, R. (2006). A default coniugate prior for variance components in generalized linear mixed models. *Bayesian Analysis*, **1**, 535–542.
- Kendall, M. G. (1946). Discussion on 'the statistical study of infectious diseases' by m. j. greenwood. *Journal of the Royal Statistical Society*, **109**, 103–105.
- Kent, J. T. (1978). Limiting behaviour of the von mises-fisher distribution. *Math. Proc. Cambridge Phil. Soc.*, **84**, 531–536.
- Kuiper, N. H. (1960). Test concerning random points on a circle. *Ned. Akad. Wet. Proc. Ser.*, **63**, 38–47.
- Laycock, P. J. (1975). Optimal design: regression models for directions. *Biometrika*, **62**, 305–311.
- Mallick, B. K. and Gelfand, A. E. (1995). Semiparametric errors-in-variables models. *Journal of Statistical Planning and Inference*, **52**, 307–321.
- Mardia, K. V. (1972). *Statistics of Directional Data*. Academic Press, New York.
- Mardia, K. V. (1976). Linear-circular correlation coefficients and rhythmometry. *Biometrika*, **63**, 403–405.
- Mardia, K. V. and Jupp, P. E. (1999). *Directional Statistics*. Wiley.
- Mardia, K. V., Goodal, C. R., and Walder, A. (1996). Distributions of projective invariants and model-based machine vision. *Advances in Applied Probability*, **28**, 641–661.

- Mardia, K. V., Taylor, C. C., and Subramaniam, G. K. (2007). Protein bioinformatics and mixtures of bivariate von mises distribution for angular data. *Biometrics*, **63**, 505–512.
- Mardia, K. V., Hughes, G., C., T. C., and Singh, H. (2008). A multivariate von mises distribution with applications to bioinformatics. *The Canadian Journal of Statistics*, **36**, 99–109.
- Matheron, G. (1962). Traite de geostatistique appliquee. *Memoires du Bureau de Recherches Geologiques et Minieres*, **N.14**, Editions Technip, Paris.
- Matheron, G. (1971b). The theory of regionalized variables and its applications. *Cahier du Centre de Morphologie Mathematique*, **N.5**, Fontainebleau, France.
- Matérn, B. (1960; reprinted 1986). *Spatial Vaiation*, 2nd ed. Springer-Verlag, Berlin.
- Rao, J. S. (1969). *Some Contributions to the Analysis of Circular Data*. Ph.D. thesis, Indian Statistical Institute, Calcutta, India.
- Ravindran, P. (2002). *Bayesian Analysis of Circular Data Using Wrapped Distribution*. Ph.D. thesis, North Carolina State University.
- Richardson, S. and Gilks, W. R. (1993). A bayesian approach to measurement error problems in epidemiology using conditional independence models. *American Journal of Epidemiology*, **6**, 430–442.
- Russell, G. S. and Levitin, D. J. (1996). An expanded table of probability values for rao’s spacing test. *Communications in Statistics - Simulation and Computation*, **24**, 879–888.
- Sherman, B. (1950). A random variable related to the spacing of sample value. *Ann. Math. Statist.*, **21**, 339–351.
- Smith, A. F. M. and Roberts, G. O. (1993). Bayesian computation via gibbs sampler and related markov chain monte carlo methods (with discussion). *Journal of the Royal Statistical Society, Ser. B*, **55**, 3–23.
- Smith, R. L. (2001). *Environmental Statistics*. Lecture Notes for CBMS Course, University of Washington.
- Stephens, M. A. (1963). Random walk on a circle. *Biometrika*, **50**, 385–390.
- Stephens, M. A. (1969e). Techniques for directional data. *Tecnical Report 150*, Departement of Statistics Stanford University.
- Stephens, M. A. (1970). Use of the kolmogorov-smirnov, cramér-von mises and related statistics without extensive tables. *Journal of the Royal Statistical Society, Ser. B*, **32**, 115–122.

- von Mises, R. (1918). Über die "ganzzahligkeit" der atomgewichte und verwandte fragen. *Phys. Z.*, **19**, 490–500.
- Waterman, T. H. (1963). The analysis of spatial orientation. *Ergeb. Biol.*, **26**, 97–117.
- Watson, G. S. (1961). Goodness-of-fit tests on a circle. *Biometrika*, **48**, 109–114.
- West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*. Springer-Verlag, New York.
- Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, **41**, 434–449.
- Wikle, C. K. (2003). Hierarchical models in environmental science. *International Statistical Review*, **71**, 181–199.
- Wikle, C. K. and Cressie, N. (1998). Hierarchical bayesian space-time models. *Environmantal and Ecological Statistics*, **5**, 117–154.